Doctoral Dissertation

Improving the safety of construction site personnel using multi-sensor data fusion

（マルチセンサデータ融合を用いた建設現場作業員の安全性向上）

CHEN TINGSONG

June 2023

Graduate School of Engineering,
Osaka University

Abstract

Construction sites are known for their inherent risks and hazards, and hazard prevention is of utmost importance in this industry. The safety of workers and those around them should always be a top priority for any construction company. Despite the advancements in safety technologies and management strategies, accidents still occur on construction sites, causing loss of life, injury, and property damage.

This dissertation explores a three-step hazard prevention approach in construction sites via multi-sensor information fusion, mainly focused on people who work on the site. Step 1 focuses on the basic workers, aiming to detect their motion to avoid potential danger. This is achieved through the use of multi-sensor units that are installed on the construction site as well as the Initial Measurement Unit (IMU) sensors attached to the workers themselves. The unit is combined with a microcomputer, depth camera, and a learning unit to achieve visual detection, the IMU sensors attached to the body of workers achieved limb signal detection. This method can achieve high accuracy in detecting specific motions of workers even beyond the suitable detecting distance of the depth camera and make a record to avoid potential hazards.

Step 2 is focused on monitoring the condition of the operators of heavy machinery to avoid fatigue and distracted operation. This is achieved through the use of a monitoring method using a Mixed Reality device, which can detect the eye gaze, head orientation and hand grasp recognition of the operator in real-time. The system set up several visual zones during the detecting process, once the operator captured any abnormal operations, such as the focus of the eyes being wrong, the head pointing position not being corrected for a long time, or the hands being off the steering wheel. The system will alert the operator visually and aurally in no time, to ensure the safety of the operation.

Step 3 aims to improve the monitoring method used in Step 2 by detecting more complicated conditions through the change of gaze, head orientation, and hand movement. By tracking the operator's eye movements, the system can detect if the operator is looking away from the machinery or not paying attention to the work area. The system can also track the orientation of the operator's head, allowing for the detection of fatigue or distracted behavior. Involuntary head nodding during extreme fatigue can be detected, as well as the large involuntary head turns when the operator is attracted. Finally, by monitoring the movement of the operator's hands, the system can detect if the operator is under normal status or not.

Since the sensors are embedded in the MR device already, adaptability and error rate control is guaranteed. When an abnormal situation is detected, the alert from the device can be operated in no time, which, in construction sites, can ensure stable and high-efficiency monitoring for the operators.

Through the implementation of this three-step approach, this dissertation aims to provide a comprehensive solution to safety management and hazard prevention in construction sites. It is hoped that the findings of this study will help improve the safety standards and practices in the construction industry and reduce the occurrence of accidents.

Preface

This dissertation is the original work by Chen Tingsong under the supervision of Professor Nobuyoshi Yabuki. Two journal articles and two international conference proceedings that relate to this dissertation have been submitted or published and are listed below.

Journal articles:

1. Chen, T., Yabuki, N. & Fukuda, T. (2023). Motion Recognition Method for Construction Workers Using Selective Depth Inspection and Optimal Inertial Measurement Unit Sensors. *CivilEng*, 2023, 4, 204-223.

2. Chen, T., Yabuki, N. & Fukuda, T. (2023). Mixed Reality-based Active Hazard Prevention System for Heavy Machinery Operators. *Automation in Construction*, (Submitted on February 26th, 2023) Under review.

International conference proceedings:

1. Chen, T., Yabuki, N., Fukuda, T. (2020). An Integrated Sensor Network Method for Safety Management of Construction Workers. *Proceedings of the International Symposium on Automation and Robotics in Construction (ISARC 2020)*. pp. 857-863, Japan, 2020.
2. Chen, T., Yabuki, N., Fukuda, T. (2021). An Active Early Warning System for Heavy Construction Vehicle Drivers based on Mixed Reality, *Proceedings of the Conference, the International Council for Research and Innovation in Building and Construction (CIB) W78 2021*, pp.793-802, Luxembourg, 2021.

Acknowledgment

Completing my doctoral thesis has been a significant milestone in my academic journey, and it would not have been possible without the support and encouragement of so many individuals.

2019 was a memorable year for me. I experienced a low point in my life during that year and doubted countless times whether I had the strength and intelligence to complete my Ph.D., a goal that seemed infinitely far away from me at that time. I also experienced a turning point in my life during that year, and I was most impressed by my choice of Yabuki Laboratory at Osaka University. It was a coincidence that I clicked on Yabuki Laboratory's homepage among so many excellent universities and research laboratories in Japan, but it was a necessity that I did not click the close button in the upper right corner. After reading Professor Yabuki's introduction and the introduction to the laboratory, I found it to be an inclusive and harmonious laboratory where the relationship between professor and student is equal and mutually beneficial. Since then, I have been convinced to continue my Ph.D. program here. I could feel the sincerity of Professor Yabuki from the beginning of my communication with him before I enrolled (Including the fact that he scheduled the interview and meetings during his business trip to Tokyo to avoid my travel), and this sincerity only increased after I enrolled and until I am about to graduate. Indeed, it was not easy to transfer from Tokyo to Osaka, but the benefits were immeasurable.

Therefore, in the beginning, from the academic view, I would like to express my deep appreciation to Professor Nobuyoshi Yabuki, thanks for his help and encouragement, and for sharing his wisdom these years. And Associate Professor Tomohiro Fukuda, for sharing his insights at each seminar and manuscript revision. As well all the members of Yabuki Lab, Mr. Jiapei Zhao, Mr. Yuehan Zhu, Miss Anqi Hu, and Miss Michael Lee, including those who graduated years ago, such as Mr. Rikuto Tanaka, Mr. Daichi Ishikawa, Mr. Hao Chen, Dr. Jiaxin Zhang, Dr. Yunqin Li, Dr. Yixi Xia, Dr. Natthapol Saovana, they support a lot since my first days in Osaka University, we had a lot of time hanging out together and enjoyed food and trips. Their knowledge, expertise, and constructive feedback were instrumental in helping me develop and refine my ideas.

From a personal point of view, I would like to thank my parents, Mr. Shengyou Chen and Mrs. Huizhen Hu, for their unwavering love and support, both emotionally and financially, which enabled me to pursue my academic goals. I am also deeply grateful to my girlfriend, Miss Qinyan Bi, for her patience, understanding, and unwavering support throughout the ups and downs of my academic journey.

I am also grateful to my friend at Osaka University, Mr. Jinfan Liu, for his wonderful cooking skills, we enjoyed many dinners in his place. Mr. Zhuolun Yang, for all the

coffee breaks under building M3, and for taking over my first-bought car, which will certainly enrich his life on the campus as well in Japan. Thanks for their support and companionship, which made my time at the university more meaningful and enjoyable.

Finally, I would like to acknowledge the contribution of my furry friends, including my naughty but cute cat “Poppy” and my fancy rats and hamsters (especially “Dasha” and “Ersha”), who provided me with a much-needed source of comfort and stress relief during the demanding academic days. I would also like to thank my friends who share my passion for cars and camping, and who provided me with a welcome distraction from the demands of my research.

Once again, I express my sincere gratitude to everyone who has supported me on my academic journey. Thank you all.

Table of contents

List of figures

List of tables

List of abbreviations

Abbreviation	Meaning
IMU	Initial Measurement Unit
MR	Mixed Reality
SDI	Selective Depth Inspection
SCAPE	Shape Completion and Animation of People
DMAS	Driving Monitoring and Assistance Systems
GradientFMM	Gradient Fast Marching Method
NCS2	Neural Compute Stick 2
VPU	Vision Processing Unit
TOPS	Trillion Operations Per Second
MRTK	Mixed Reality Toolkit
BIM	Building Information Modeling

Chapter 1

Introduction

1.1 Background

The concept of “Smart City” has been widely known by people around the world for many years. Its main purpose is to use information technology to help to improve city ser-vices. Currently, the smart city concept is being applied in transportation, citizen management, and urban resource allocation, all with good performance compared with traditional methods. Along with the rise of the smart city, the concept of “smart construction” (Xu & Lu, 2018) has also been proposed recently, in which many well-applied methods from other fields are exported to the construction area. However, owing to the differences in management mode and implementation, these approaches have not performed well. Thus, there is much room for improvement in the development of smart construction.

This lackluster performance is especially concerning in light of the construction industry field currently holding the worst record for safety compared with other industries: approximately 88% of workplace incidents in the construction industry are caused by unsafe behaviors (Shin et al., 2015). For example, in Japan, the average number of fatalities in construction accidents is over 300 people per year (*Statista,* n.d.), and this performance has not improved well during the last 20 years (*Kensaibou,* n.d.). In comparison with other developed countries, accident monitoring efficiency in Japan is the main reason that injured workers cannot be identified and rescued in time.

After recent years of development and the wide application of deep learning, image processing has greatly improved the accuracy of human motion detection. However, single visual sensor detection still has its unreliability, it is easily affected by changes in the surrounding environment and the location of the measured object.

Meanwhile, wearable devices have become increasingly popular recently. Some simple devices such as watches can perform basic body path or state detection. However, if these devices are the only angle of detection, they lack reliability. Complex special

clothing with hundreds of detection points can perform extremely detailed tracking. However, because of its inconvenience, worker resistance, and high cost, it is not suitable for use in ordinary construction environments.

In the meantime, another necessary component in a construction site should not be ignored: The introduction of heavy vehicles and machinery has benefited the construction industry because they have made many tasks easier and reduced the workload and fatigue of workers. However, everything has advantages and disadvantages, and heavy vehicles are no exception. Due to their inherent objective limitations, the potential for injury to workers is not outweighed by the convenience that heavy vehicles provide. On many construction sites, workers are forced to share the worksite with heavy vehicles, increasing the chance of accidents due to the bulkiness of the vehicles. In many fatal accidents involving heavy vehicles, the cause of the accident is often improper operation by the driver, but it is other workers who tend to sustain the most serious injuries or even fatalities.

According to a previous study (Olorunfemi et al., 2018), Mixed Reality (MR) is a visualization environment that combines the virtual world and the real world, creating a virtual space in which digital and physical objects coexist and interact in real-time, giving users have a comprehensive holographic experience.

By superimposing virtual digital content in the real world, users can instantaneously experience enhanced interactions with the real world. MR has moved from the experimental stage to practical applications in fields such as medicine and healthcare, education, entertainment, and industrial maintenance.

1.2 Motivation for the research

In construction sites, the main causes of worker injury and death are related to health issues such as heat stroke, physical injuries from falling, contacting, and mental injuries. Some of the injuries are caused by accidents, and some of them are caused by unsupervised unsafe acts taken by workers as a result of cost limitations, time pressure, and other reasons. Normal monitoring systems, such as web cameras, require manual operation and are inefficient because of human neglect, visual obstacles, and other factors. A more efficient and accurate safety management and monitoring method is needed.

Camera-based monitoring methods are widely used and researched around the world because of their many benefits, such as low cost and ease of assembly. However, unavoidable disadvantages are more numerous, such as low accuracy when it is too far away from the visual sensor or in poor lighting conditions. Some approaches concentrate on changing the RGB frame to an RGB-D frame by adding depth to pictures with machines such as Kinect (Lun & Zhao, 2015) and RealSense (Rabbani et al., 2020).

RGB-D is more accurate for detecting humans and objects compared with RGB pixel cropping, and it also works well even under poor lighting conditions. However, owing to the limitations of working distance, reflective surfaces, and relative surface angles, depth maps in RGB-D frames always contain significant holes and noise, and these errors limit the practical use of RGB-D frames in real ap-plications. Thus, depth maps for filling holes and removing noise are necessary steps in depth camera-based monitoring systems. Because of the high error rate of depth measurement over wide areas and long distances, depth scanning of large areas also increases the calculation demands. Therefore, an approach that can effectively achieve an in-depth analysis of designated areas is needed.

Monitoring methods based on inertial measurement unit (IMU) sensors are also gaining attention in recent years because of their clear benefits compared with other methods, such as those relying on visual cameras. IMU sensors are non-intrusive, light-weight, and portable measuring devices that, when attached to a subject, can overcome the sensor viewpoint to detect activities in a non-hindering manner (Ann & Theng, 2014; Dehzangi et al., 2017; Cismas et al., 2017). After preprocessing of the motion for recognition, discriminative features are then derived from time and/or frequency domain representations of the motion signals (Preece, Goulermas, Kenney, & Howard, 2009) and used for activity classification (Preece, Goulermas, Kenney, Howard, et al., 2009). Although there are many benefits of IMU sensors, there are also some disadvantages. First, their output is not intuitively understood and not amenable to manual rechecks. Second, model complexity is hard to control, especially when precise motion capture is needed. Generally, construction workers stay outdoors, and the changing environment there makes them more resistant to accepting a huge number of wearable devices, such as some complex special clothing with hundreds of detection points on it. Therefore, reducing the burden on workers while stabilizing detection accuracy is an issue that needs to be resolved.

The element of danger on a construction site is much more than simply the dangerous behavior of the workers themselves. Heavy machinery on the site that is meant to help workers in order to improve efficiency can also cause significant damage if not operated properly. For example, according to the record in the United States, on average more than 20,000 workers are injured as a direct result of traffic-related accidents during construction activity each year. In 2016, more than 25,000 workplace accidents occurred in Texas, resulting in the fatalities of 18 people (*Steven M. Lee*, n.d.). When a driver or operator is distracted or inattentive due to insufficient attention to the surrounding environment, it is possible to knock down or even run over road construction workers or deviate from the established route and collide with other vehicles. Similarly, some drivers who are not safety conscious on public roads also pose a threat to workers when passing through construction areas.

In the Japanese construction industry as well, such incidents are not uncommon. According to a questionnaire survey of more than 100 Japanese construction companies, more than half reported having vehicle collision accidents during the construction process.

Previous reports have shown that the main causes of major fatal accidents involving heavy vehicles include inattention when overtaking or reversing, collisions between vehicles, and collisions between vehicles and equipment or workers.

It has been reported that 85% of vehicle control depends on the state of the driver (Khan & Lee, 2019), and thus, when considering a safety warning system for large vehicles, more focus should be placed on the driver than on the vehicle itself. Figure 1 shows a truck that attempted to turn while speeding and failed, while Figure 2 shows that the driver's inattentive driving caused the vehicle to enter the opposite lane, thereby causing an accident.

Figure 1 Simulated accident of a heavy vehicle turning while speeding (Reprinted with the permission of the copyright owner).

Figure 2 Simulated accident caused by inattentive driving (Reprinted with the permission of the copyright owner).

Most of the studies on hazard prevention methods for drivers or operators of machines are based on prevention, mainly in the form of extensive training and safety education for the operator to increase his safety awareness, but the reliability of these methods also depends largely on the attitude and state of the operator. The efficiency of these methods becomes low when there are some subjective uncontrollable situations.

In the meantime, according to a previous study (*Mixed Reality,* n.d.), the use of MR in various industries is also starting to increase gradually. More and more construction companies and organizations are also conducting research on it because of its virtual-reality nature which is highly compatible with the construction industry. MR is a visualization environment that combines the virtual world and the real world, creating a virtual space in which digital and physical objects coexist and interact in real-time, giving users a comprehensive holographic experience.

By superimposing virtual digital content in the real world, users can instantaneously experience enhanced interactions with the real world. MR has moved from the experimental stage to practical applications in fields such as medicine and healthcare, education, entertainment, and industrial maintenance.

1.3 Research questions and objectives

A construction site is a complex environment that can include many corners, blind spots, and isolated spaces. When using only one type of sensor (such as a vision sensor) for worker hazard identification, the accuracy of the sensor can vary greatly due to lighting, camera angle, etc. For example, once an object or worker is outside the optimal

recognition area of a vision sensor, other types of sensors must be used to assist in recognition to ensure accuracy.

To address the issues discussed above, an advanced worker hazard prevention method based on a multi-sensor network is proposed. The approach aims to use worker data obtained from different types of sensors to more accurately detect human motion and movement, including RGB cameras, depth cameras, and IMUs, and to combine them cooperatively to improve the accuracy of detecting specific worker movements and to improve the efficiency of accident warning and injury rescue.

To address this research gap, the purpose of this study is to examine the application of deep learning in removing unwanted features from repetitive infrastructure images, an application that is both laborious and time-consuming. The impact of removing unwanted features on the process of moving from motion to the structure is also investigated. First, a new method called Selective Depth Inspection (SDI) is proposed. This method adds preprocessing and image aging assistance to the common depth map optimization, thus significantly improving computational efficiency and accuracy. Second, a multi-sensor-based construction site motion recognition system is proposed, which combines different kinds of signals to analyze and correct workers' motions on the construction site to improve the accuracy and efficiency of detecting specific body motions on the construction site.

In hazard detection for heavy machinery operators, the possibility of accidents increases dramatically if the operator himself, who dominates the operation, shows some abnormal behavior or phenomenon. A more stable monitoring method is necessary to ensure that the operator can be alerted in any state (awake or distracted, etc.), rather than relying on the operator's subjective awareness.

One of the technical features of MR is its ability to use Building Information Modeling (BIM) to display and manage holograms, which makes it highly adaptable to the construction field. Another important research direction is the use of remote monitoring to enhance presence. For example, a previous study (Olorunfemi et al., 2018) used holographic MR technology running on Microsoft HoloLens (*Microsoft HoloLens*, n.d.) to enable visual interaction and remote collaboration to discuss site risks at construction sites.

Although most MR-based studies have focused on helping users perceive and interact with their environment through MR devices, few applications have utilized the internal cameras and sensors of MR devices to monitor the users themselves. Therefore, an active driving condition monitoring and early warning system based on MR is proposed to focus on real-time monitoring using individual sensors in order to analyze the current state of the wearer (heavy machinery operator) to correct problems and prevent accidents in a timely manner (T. Chen et al., 2021).

1.4 Overview

Figure 3 shows the overview of this research and the dissertation is organized as follows.

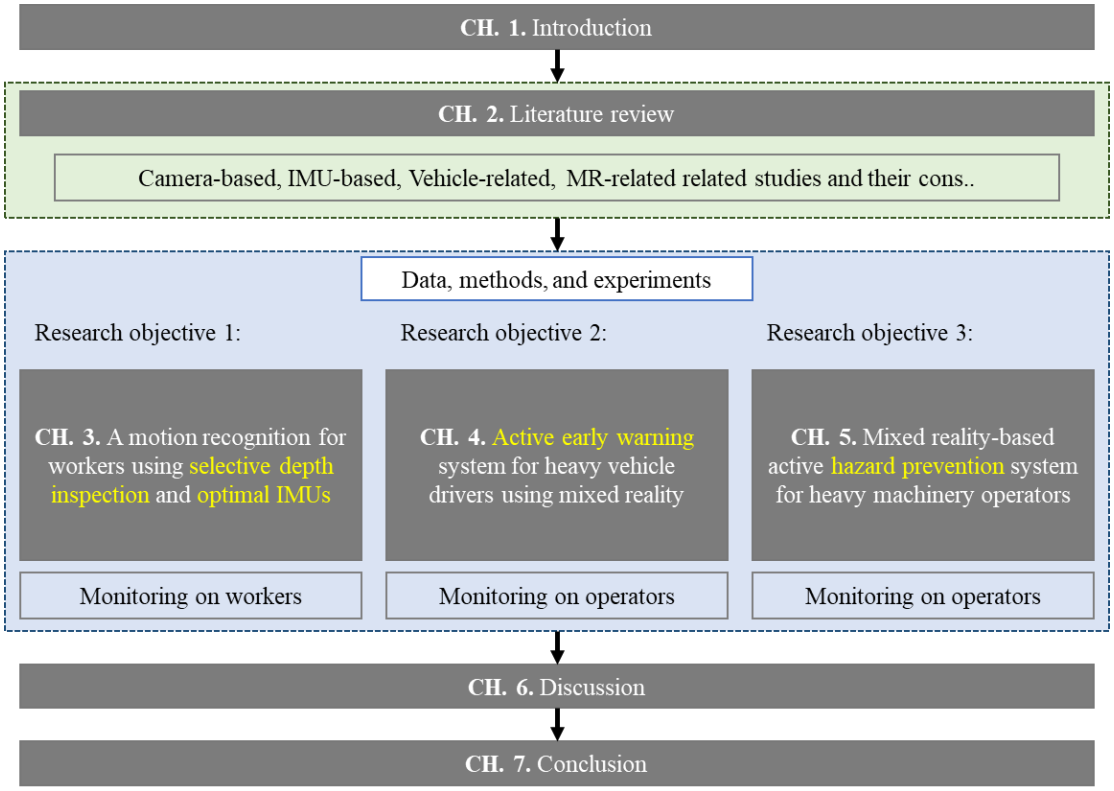


Figure 3 Overview of this research.

Chapter 1 Introduction

This chapter introduces the background information about this study and names the motivation for conducting this study. In the third subsection, the research questions, previous shortcomings, and the purpose of this study are highlighted.

Chapter 2 Literature review

This chapter briefly describes the directions and achievements of the research, including the first experiment-based research, such as the drawbacks and shortcomings of single-sensor research, and the second and third experiment-based research, such as the content of driver control assistance systems, and the application of MR in the construction industry.

Chapter 3 Motion recognition for workers using selective depth inspection and optimal IMUs

In this chapter, a worker motion recognition method based on depth cameras and IMU sensors is proposed to maintain high motion recognition accuracy beyond the optimal depth detection distance. By combining IMU sensors with depth cameras, this research can identify workers' movements more accurately and thus prevent and avoid potential safety risks in time.

Chapter 4 Active early warning system for heavy vehicle drivers using mixed reality

In this chapter, an MR-based active safety monitoring solution for operators of large machines at construction sites is proposed. By wearing HoloLens 2 and monitoring the operator's eye gaze area, head pointing area and hand movements in real-time, the operator can be alerted by auditory and visual signals in case of distraction or fatigue to avoid accidents.

Chapter 5 Mixed reality-based active hazard prevention system for heavy machinery operators

This chapter is an improvement and enhancement based on the content of Chapter 4. After implementing basic eye, head, and hand monitoring, a series of operator movements are monitored and analyzed to further ensure the operator's status and prevent accidents. This technology allows for more comprehensive safety monitoring by monitoring specific movements such as frequent head nodding, disoriented eyes, large head bobbing, and hands off the steering wheel when the operator is tired. Meanwhile, this research considered the influence of indoor and outdoor environments in the experiment and used multiple random sampling in the result analysis to ensure objectivity.

Chapter 6 Discussion

This is the discussion section where the findings of the above three experiments will be briefly explained, and the results obtained will be analyzed and evaluated.

Chapter 7 Conclusions

This is the concluding chapter, which summarizes the study, describes the significance and conclusions, discusses the limitations and shortcomings of the study, and briefly introduces future work.

Chapter 2

Literature review

This chapter presents the related work of this dissertation. It is organized as follows. Sections 2.1 and 2.2 present research related to motion recognition for construction site workers, 2.1 presents a vision-based human modeling approach, and 2.2 presents an IMU sensor-based human modeling approach. Sections 2.3 and 2.4 present research related to safety management for large machine operators on construction sites, 2.3 presents the main elements of a driving safety assistance system, and 2.4 presents the mainstream MR or virtual reality (VR) applications and development in the construction industry.

2.1 Visual-based human modeling methods

Among the studies on action recognition based on visual sensors, pure vision-based studies occupy the majority due to their very low cost and simple entry channels, while some studies use more advanced depth cameras in an attempt to obtain more stable and valid information through an additional dimension.

2.1.1 RGB camera-based human modeling

The use of RGB camera-based motion recognition is a cost-effective solution for many applications and can be easily deployed in various environments. However, it is limited by its sensitivity to lighting conditions and is susceptible to interference (Guo & Dai, 2018).

Commercial camera-based human detection systems often require the use of markers or multiple camera setups (Sarafianos et al., 2016), which can be inconvenient for users. Markerless approaches have been developed using multiple cameras, but these methods typically require offline processing to achieve accurate results (Ballan et al., 2012; Huang et al., 2017)

Some real-time markerless approaches have been proposed, such as those that combine a skeletal model with image data or generative and discriminative methods (Elhayek et

al., 2017; Rhodin et al., 2015). However, these approaches may still require well-calibrated cameras and are not suitable for mobile scenarios.

Despite these limitations, ongoing research is focused on developing more robust and efficient markerless approaches for human motion detection and recognition. These advancements will enable the deployment of cost-effective and easy-to-use motion recognition systems in a wider range of scenarios, including mobile and dynamic environments.

2.1.2 Depth camera-based depth map restoration and human modeling

Most motion recognition based on depth cameras is performed at a short distance, so its range of practical application is relatively narrow and cannot be easily expanded (Amine Elforaici et al., 2018; C. Chen et al., 2015). As for regular cameras, RGB-based depth prediction normally relies on a large body of literature and is trained with ground truth data only (Bo Li et al., 2015; Eigen & Fergus, 2015; Liu et al., 2015; Zhang et al., 2015; Zhou et al., 2017)

In the case of depth cameras, various approaches have been proposed to restore depth maps with Kinect, including two main types: filtering-based methods and reconstruction-based methods. Filtering-based methods use different kinds of filters to restore captured depth maps. For example, a median filter has been proposed to recursively fill holes in a depth map, but the sharpness of the edges is often too blurred (Lai et al., 2011). To maintain sharp edges, a joint bilateral filter has been applied iteratively to the depth map (Camplani & Salgado, 2012).

(Matyunin et al., 2011) considered using temporal information to restore the depth map, but this approach incurs a delay because it uses multiple consecutive frames for the restoration. Methods that are based on reconstruction use image inpainting to fill missing values in depth maps. For example, the fast-marching method (FMM) has been proposed for depth value reconstruction (Telea, 2004). A texture-assisted approach has also been proposed, in which texture edge information is extracted to aid the restoration of depth values (Miao et al., 2012).

However, these methods may have difficulty filling large holes in in-depth maps, resulting in unsatisfactory results as illustrated in Figure 4. While these methods can remove noise and fill small holes in in-depth maps, more robust techniques are needed to address larger gaps in in-depth information.

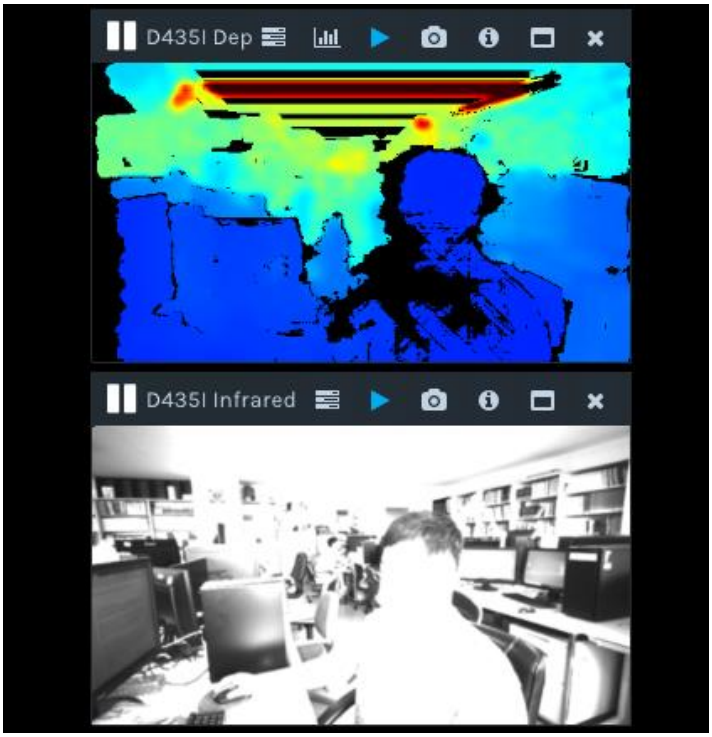


Figure 4 Depth map (top) and infrared frame.

To address human modeling, researchers have proposed a method named Shape Completion and Animation of People (SCAPE) (Anguelov et al., n.d.), which is a data-driven approach to 3D human model building from both shape and pose aspects. The SCAPE model can obtain observation data with a high-resolution depth image from a single viewpoint.

Building upon the SCAPE model, (Weiss et al., 2011) combined low-resolution scans and views of a person from different angles to construct an accurate human 3D model.

(L. Liao et al., 2017) proposed a monocular depth camera 3D human modeling approach based on referring to previous human body pose and shape approaches. Their method relies on a single-depth camera to capture human motion and construct a 3D human model.

2.2 IMU-based human modeling methods

One study (Roetenberg et al., 2013) used 17 Inertial Measurement Units (IMUs) with the function of magnetometers, gyroscopes, and 3D accelerometers, which were fused by a Kalman filter. Although these IMUs can define the full pose of the subject in standard skeletal models, they are intrusive and difficult to reproduce due to their high number. Additionally, problems such as long setup times and placing sensors in the wrong position are common.

A study (by Marcard et al., 2017) achieved accurate 3D poses using only six IMUs. They placed synthetic IMU sensors on a skinned multi-person linear body model in a generic way and solved for the sequence of body poses that matched the observed actual measured sequence by optimizing the entire sequence (Loper et al., 2015). However, this approach relies on computationally expensive offline optimization, which makes it difficult to reproduce.

Wearing a large number of sensors can be uncomfortable and impractical for long-term outdoor work, and it also increases the computational burden on the system (Rogez & Schmid, 2016). Therefore, researchers have been exploring ways to achieve accurate motion detection using a smaller number of IMU sensors and less complex algorithms. For instance, (Caputo et al., 2018) proposed a body pose recognition system that used six IMU sensors placed on the experimenter's arms, neck, and waist to achieve high precision in detecting posture angles for fixed movements. However, the system was limited to recognizing only two movements, standing and bending.

In contrast, a study (Hirota & Murakami, 2016) implemented motion command analysis for an electric wheeled walker using only one IMU sensor. The system analyzed acceleration and triggered emergency behaviors such as electric wheel braking to ensure the carrier's safety. However, the study had significant limitations in terms of the diversity of motion and the applicability of the system.

Another study (Bangaru et al., 2022) proposed a fatigue testing framework based on the construction industry that used a single IMU and forearm electromyography to evaluate physiological indicators. However, the application was relatively narrow since it focused only on fatigue testing. In comparison, a study (Mekruksavanich & Jitpattanakul, 2023) identified construction workers' activities using information from three IMU sensors, achieving high recognition accuracy. However, the placement of sensors on only one arm limited the system's ability to recognize complex limb movements.

2.3 Driving monitoring and assistance system

2.3.1 Vehicle safety-related research

The growing emphasis on safety in the construction industry has spurred a significant amount of research on construction vehicle safety, with a particular focus on the development and implementation of driving monitoring and assistance systems (DMAS). These systems are designed to enhance the safety of construction vehicles by assisting drivers with various tasks and monitoring their behavior to prevent accidents and improve overall performance. There are three main types, which are described in the following sections.

2.3.2 Driver-focused research and studies

Driver distraction is one of the leading causes of car-related accidents, accounting for 50% of all driving accidents (Klauer et al., 2006). Driver distractions include mainly cognitive, olfactory, biomechanical, auditory, gustatory, and visual distractions.

Previous studies (Angell et al., 2006; Y. Liao et al., 2018) have reported that research and analysis on gaze can differentiate between focused and distracted driving, thereby providing relative degrees of cognitive engagement. Another study (Hayhoe, 2004) reported that cognitive workload, distraction, and eye movements (e.g., staring, chasing, scanning) are correlated. Furthermore, as the cognitive workload increases, the driver's head movement tends to increase as well. The increase in head movements can be explained as a compensatory action for the driver to improve their field of view. According to one study (Miyaji et al., 2009), a driver's cognitive dispersion state can be properly detected by the standard deviation of head and eye movements. Meanwhile, another study (Liang & Lee, 2010) reported that frequently looking at distant objects increases the dispersion of vision; the frequency of blinking increases with the dispersion of cognition, while the concentration of gaze and the decrease in saccade are indications that the driver is visually and cognitively distracted.

The previous study (Khan & Lee, 2019) suggested that about 25%–35% of car-related accidents are caused by fatigued driving. Driver fatigue is divided into three main types: drowsiness caused by mental and central nervous system fatigue, which is the most dangerous type; general fatigue, such as the tiredness that happens after a long day at work; and local fatigue, such as skeletal and muscle fatigue caused by prolonged sitting.

Fatigue also substantially impacts driving performance, with some effects being directly related to the driver's physical state. These effects include increased blinking frequency caused by yawning; burning eyes; difficulty maintaining a firm grip on the steering wheel; erratic eye movements; long response times; and increased nodding (Eskandarian et al., 2008; Z. Li et al., 2017; Mandal et al., 2017). These studies highlight the serious consequences that can result if the driver becomes unconscious while driving.

2.3.3 Vehicle-focused research and studies

Driver-focused safety systems are designed to monitor the physiological state of the driver, such as fatigue, distraction, and drowsiness, to enhance safety. On the other hand, vehicle-based safety systems take into account factors like driving style and environmental conditions to provide a personalized driving experience. This personalized approach takes into account individual differences in personality, age, and behavior and can adapt to changing road conditions. However, a significant disadvantage of a highly personalized system is that it can lead to major errors if the system becomes overly reliant on certain assumptions or data. Additionally, the same

individual may perform differently at different times due to varying physical, emotional, or mental states. Therefore, such systems need to strike a balance between personalization and generalization to ensure the highest level of safety possible. By combining both driver-focused and vehicle-based safety systems, it is possible to create a comprehensive safety framework that accounts for both environmental and human factors, leading to a safer and more efficient driving experience for everyone on the road (Filev et al., 2009; G. Li et al., 2015).

2.3.4 Driving environment-focused research and studies

DMAS is crucial for enhancing driving safety and has been widely implemented in real-world driving scenarios. Many vehicles on the road today are equipped with basic safety functions, while some electric vehicle brands may have more advanced systems. DMAS improves drivers' attention by detecting surrounding vehicles and pedestrians, helping to prevent collisions. These systems rely on a variety of sensors, including passive video and audio sensors (J. Kim et al., 2012; Mizumachi et al., 2014), as well as active sensors such as radar and lidar (Cho & Tseng, 2013; Nashashibi & Bargeton, 2008), to capture information about the vehicle's status and the surrounding environment. With the aid of these sensors, most accidents can be avoided when the driver is fully in control. However, if the driver becomes unconscious, the accident rate can be significantly higher. Therefore, it is important to continue researching and developing new technologies that can detect and respond to unconscious drivers to further enhance driving safety.

2.4 MR-related research in the construction field

By analyzing MR research related to the construction industry, The truth that most studies focused on education and skills training, on-site environmental monitoring, and pre-construction planning is founded.

2.4.1 Education and skills training

MR (MR) has numerous applications across industries, and one of the most common uses is in providing training and education for students, employees, and managers. The unique properties of MR make it possible to provide training for hazardous jobs and simulate work environments without exposing users to actual risks (Jeelani et al., 2017). For example, a previous study (H. Li et al., 2012) developed a multi-user training program using a Nintendo Wii game controller, which allowed trainees to learn crane dismantling skills and practice them in a virtual environment. Another study comparing students who learn from an MR training program with those who learn from a skilled trainer showed that better training results can be achieved in an MR learning environment.

Several studies have highlighted the potential for MR in addressing the shortage of skilled workers in the labor market, specifically in the construction industry (Azhar et al., 2018; Elrawi, 2017; Lu et al., 2015; Wang & Dunston, 2013). Another study (Ogunseiju et al., 2022) explored ways to provide virtual education to construction students while helping them understand the precision and significance of construction activities. Communication has been identified as a crucial factor in improving safety performance in construction, as emphasized in various studies (Alsamadani et al., 2013; Christian et al., 2009; Haslam et al., 2005). Additionally, a study developed a system of key safety and quality performance indicators to evaluate potential benefits (Shohet et al., 2019).

Therefore, MR technology has the potential to revolutionize education and training across industries, providing realistic simulations and a safe environment for learning and skill-building. As technology continues to evolve, it is likely to become an increasingly important tool in workforce development and safety training.

However, the use of MR in safety training and education can also have limitations, such as the cost of implementing the technology and the need for specialized expertise to create virtual environments. Furthermore, while MR can provide a realistic simulation of hazardous situations, it may not fully capture the stress and pressure that workers may experience in real-world scenarios. Therefore, it is important to supplement MR training with real-world experience and provide ongoing support and feedback to ensure that workers can effectively apply what they have learned in both simulated and actual situations. Ultimately, a comprehensive approach that combines MR training with real-world experience and real-time safety systems can help improve safety performance in high-risk industries.

2.4.2 On-site environment monitoring

On-site monitoring often connects MR systems with location trackers and shares valuable safety information in real-time. In one study (K. Kim et al., 2017), researchers proposed a Google Glass (*Google Glass,* n.d.)-based system for tracking workers in real-time in order to expand their field of view and visually warn them when they approach a dangerous area or machine, and at the same time, the operator of the heavy machinery will receive an alert that someone is approaching. Another study (H. Li et al., 2015) focused on VR-based real-time tracking and training in which hazardous areas in the virtual environment are marked and radio frequency identifiers are attached to the workers' helmets. Whenever workers enter a hazardous area or approach heavy machinery, the helmet emits an audible warning alerting them to move away. This approach also involves a virtual environment on site called the "virtual construction simulation system," which can be monitored by managers and safety professionals. The researchers concluded that despite the limited realism provided by real-time visualization, augmented reality (AR) and MR will be the mainstay of construction

monitoring in the future. In another study (Kun et al., 2018), the participants used MR equipment and made Skype calls while driving to test whether the MR equipment negatively affected driver judgment. The experiment was performed in an isolated environment and carried out under strict safety standards. In yet another study (Wu et al., 2022), digital twins, deep learning, and MR were integrated into a real-time visual warning system. Finally, another study (Dai et al., 2021) reported that MR could be used to communicate and interact visually on the construction site and discussed the potential for preventing on-site accidents.

These results demonstrate the practicality of using MR devices or similar equipment when performing operations outdoor. At the same time, the sensors installed in the equipment allow the identification system to more accurately determine the location and status of personnel, thereby facilitating more comprehensive protection.

2.4.3 Pre and post-construction planning

Preconstruction safety planning is a crucial step that should be taken before starting any construction project. It can help to prevent potential hazards and ensure the smooth progress of the project by ensuring that proper design specifications are followed. One promising method for carrying out preconstruction safety planning is through the use of MR technology.

One study proposed a VR-based 4D simulation system (Boton, 2018) to support the analysis of designers and construction managers. This system allows them to identify potential safety hazards and develop strategies to mitigate them before construction begins. Another study (Malekitabar et al., 2016) investigated the use of Building Information Modeling (BIM) to automatically identify construction safety issues and related risks.

Therefore, preconstruction simulations are essential for ensuring the safety and efficiency of the actual work and the efficient use of resources by reducing overturns and preventing false starts. MR-based simulations are expected to become a key preconstruction safety planning method in the construction industry in the future.

In addition to preconstruction safety planning, MR can also be applied to post-construction safety assessment and management. One study (Khurshid et al., 2023) proposed a system for monitoring and assessing the safety of building structures using a combination of MR and Internet of Things (IoT) technology. The system uses sensors to collect data on structural deformations and sends this information to a cloud-based MR platform, which can generate 3D visualizations of the building's structural integrity in real-time. This approach provides a more comprehensive and accurate assessment of a building's safety compared to traditional manual inspections.

Overall, MR has significant potential to improve safety in the construction industry, from preconstruction planning to post-construction assessment and management. By providing workers and managers with real-time information and feedback, MR systems can help to prevent accidents and injuries and improve the overall efficiency and effectiveness of construction projects.

2.5 Summary

In human motion, recognition approaches that are constrained to use a single sensor, certain hardware limitations may arise, such as utilizing a depth sensor, which may not provide accurate readings for targets that are beyond optimal range. Likewise, if only IMU sensors are used, the interference of the environment can lead to significant integration errors. These issues are particularly important in the complex environment of construction sites. Most studies using a single sensor are based on many conditions, whereas studies using multi-sensor fusion are necessary for complex environments and actions.

Meanwhile, based on the research described above, it is clear that conventional driver-based safety monitoring and early warning methods involve the measurement of biological indicators, which can be difficult to implement in everyday use in general environments. Many safety-related studies on vehicles and drivers have focused on the provision of information. For example, for vehicles with poor visibility, a large number of sensors are utilized to enhance peripheral vision and provide distance information so that drivers can actively assess their safety levels. Workers who operate large machinery generally have ample on-the-job experience. The main reason they have accidents is that their subjectivity is disturbed, keeping them from receiving or evaluating information in time, such as when they are fatigued or distracted. Therefore, safety research on drivers should focus on monitoring and analyzing the status of the driver while simultaneously providing warnings and reminders in a timely manner. Meanwhile, the complexity of the processing equipment makes it difficult to apply in real-world environments. On the other hand, MR-based safety research in the construction industry is macroscopic and tends to focus more on prior training, advanced layouts and hazard prevention, and remote human monitoring, but the response to emergencies remains insufficient.

Chapter 3

A motion recognition for workers using selective depth inspection and optimal IMUs

The construction industry has long been plagued by poor safety records, with accidents resulting in worker injury being all too common. With approximately 88% of accidents resulting in injury, there is a pressing need for improved hazard prevention methods to keep workers safe on construction sites. In recent years, deep learning and image processing have shown great potential for improving human motion detection accuracy, but equipment limitations have made it difficult to effectively address depth-related problems. Wearable devices have become increasingly popular, but the variable outdoor environment in which construction workers operate presents significant challenges to their widespread adoption.

To address these challenges, an integrated sensor fusion method is proposed for hazard prevention on construction sites. The approach combines different kinds of signals to analyze and correct the movement of workers on the site, improving the detection accuracy and efficiency of specific body motions. The proposed approach incorporates a new method called selective depth inspection (SDI), which adds preprocessing and imaging assistance to an ordinary depth map optimization, significantly improving calculation efficiency and accuracy.

The SDI method combines depth cameras and inertial measurement unit (IMU) sensors to achieve more accurate motion recognition, even beyond the suitable detecting distance of depth cameras. This combination allows for the detection of body movements that are not easily captured by depth cameras alone, making it a valuable tool for hazard prevention on construction sites. The proposed approach can reduce the burden on workers while stabilizing detection accuracy, making it an attractive solution for construction site safety.

The multi-sensor-based motion recognition system for construction sites includes several key components. First, the system uses a depth camera to capture images of workers on the site. The SDI method is then used to optimize the depth maps and

improve calculation accuracy. Second, IMU sensors are attached to workers to capture data on their movements. This data is then analyzed in conjunction with the depth images to improve the detection accuracy of specific body motions.

3.1 Methodology

3.1.1 Overview

The proposed method for hazard prevention of construction workers is an integrated sensor network approach, which utilizes a depth camera and IMU sensors to collect data from workers and construct a human model to analyze their motions and gestures. As shown in Figure 5, this approach involves a two-step process: visual-based motion recognition and IMU-based motion recognition. The visual-based motion recognition step uses image-processing-based real-time monitoring as a preprocessing step to enhance the accuracy of depth optimization. On the other hand, the IMU-based motion recognition step minimizes the number of wearable devices depending on the application environment while maintaining the detection accuracy of basic and relatively complex actions.

One of the main advantages of this method is that it focuses on multi-sensor cooperation, which helps reduce errors caused by sensor defects, increase detection accuracy, and improve efficiency. Additionally, the use of a depth camera and IMU sensors allows for more accurate and comprehensive monitoring of workers' motions and gestures, which can help prevent potential hazards in the workplace.

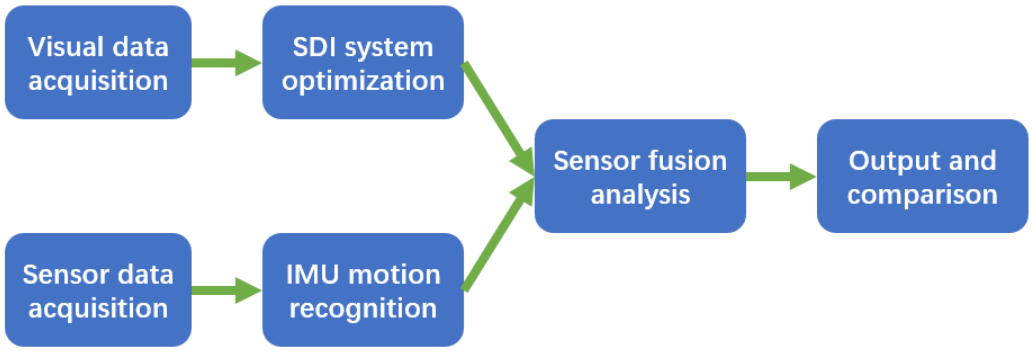


Figure 5 Overview of the proposed methodology.

However, it is important to note that the implementation of this approach may require additional resources, such as specialized hardware and software, which can increase the cost and complexity of the system. Moreover, there may be limitations to the range and accuracy of the sensors used, which can affect the detection and recognition of motions and gestures. Therefore, further research and development may be necessary to optimize the effectiveness and practicality of this method for hazard prevention in the construction industry.

3.1.2 Human recognition via SDI

The proposed selective depth inspection (SDI) method addresses the challenges associated with depth detection in complex environments, such as construction sites, by dividing the process into two steps. The first step utilizes image processing to identify the monitoring area and check for the presence of humans. This step helps to reduce the computational burden and avoids performing depth recognition and optimization on non-human objects, thereby improving efficiency. In the second step, depth recognition and optimization are performed on the selected areas, which reduces the recognition errors caused by specific actions and increases the effective recognition distance. This approach is particularly useful for situations where the object of interest is located at a distance from the camera, as it helps to reduce the high error rates that may occur in typical depth detection over long distances. By effectively combining image processing and depth optimization, the proposed SDI method can enhance the accuracy and efficiency of human motion recognition in complex environments, making it a valuable tool for hazard prevention in construction sites.

The proposed preprocessing stage involves utilizing various image recognition methods to identify the selected area. Two methods are used: human recognition based on TensorFlow Lite (*TensorFlow Lite*, n.d.) and skeleton recognition using PoseNet (*PoseNet*, n.d.), both of which are based on TensorFlow Lite. The human recognition method quickly identifies potential human shapes in real-time through analysis, with varying accuracy based on the training sets used. The skeleton recognition method performs rough bone recognition through an RGB camera to enhance the accuracy of depth analysis and optimization. To ensure recognition efficiency, a microcomputer is used as the terminal in the proposed method. Although the models used, such as TensorFlow Lite and PoseNet, are slightly outdated, they still meet the experimental requirements of a low-consumption terminal and provide an acceptable recognition fluency.

When a human signal appears in the monitoring range, the RGB camera will detect the body frame, and only the depth data inside the selected area will be collected. The raw depth information collected may cause too many holes and defects owing to factors such as distance and environmental interference. To fix the hole problem and other interference, depth information optimization is needed at this time. This optimization

can fill in the missing depth information at a long distance to effectively increase the effective distance of depth motion recognition.

A study (Yin et al., 2019) proposed a two-stage stacked hourglass network based on a previous study (Varol et al., 2017) to obtain high-quality results for human depth prediction. Instead of using RGB images directly, this approach uses RGB images and human part segmentation together to predict human depth. It consists of convolution layers, a part-segmentation module, and a depth prediction module. First, the RGB image input goes through the convolution layer and is converted into heat maps, after which it enters the part-segmentation module. Then, the heat maps are converted into human part-segmentation results, and these heat maps are summed as the input of the following depth prediction module with the features of previous layers. Finally, human depth prediction results are output.

Algorithm 1: Gradient Fast-Marching Method (GradientFMM)

```
1.  Procedure GradientFMM (depthmap)
2.  Known ← all pixels with known values in depthmap
3.  Unknown ← all unknown pixels adjacent to Known in depthmap
4.  insert all pixels in Unknown into min-heap
5.  while Unknown not empty do
6.     p ← root of min-heap
7.     calculate p values using depth value equation
8.     add p to Known
9.     remove p from Unknown
10.    perform down heap
11.    for each neighbor q of A do
12.      if q not in Known and Unknown then
13.        add q to Unknown
14.        perform up heap
15.      end if
16.    end for
17.  end while
18.  return Known
19. end procedure
```

Figure 6 Gradient FMM algorithm.

Figure 6 above is an algorithm called Gradient Fast Marching Method (GradientFMM) (Yin et al., 2019), and it propagates the depth from known pixels to unknown pixels. After the process, every pixel in the unknown region of a depth map will be assigned a depth value. In this study, to extend the detectable distance of the selected area, the GradientFMM algorithm is applied for depth information optimization.

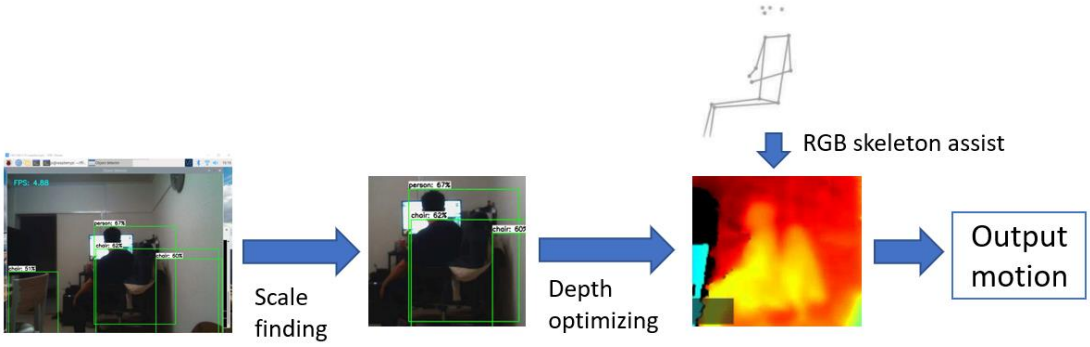


Figure 7 SDI method process.

The depth camera produces images with a resolution of 848 × 480 pixels at a framerate of 30 frames per second. As shown in Figure 7, in the optimization process, this research first applies the GradientFMM algorithm to analyze each frame to fix undetected points of selected human areas, and then the area is considered as 3D coordinates by cooperation with the skeleton detection processing from the previous step. Finally, the depth information as a three-dimensional coordinate of each point of interest from the possible hu-man shape can be obtained. The points of interest include but are not limited to the hands, elbows, head, waist, knees, and feet.

3.1.3 Portable computing terminal

Some hazard prevention schemes are based on sensors placed on helmets, and the information collected by sensors on helmets can be analyzed only after workers end their shifts and remove their helmets. This kind of analysis method is relatively inefficient and cannot provide timely accident alerts.

In this study, as shown in Figure 8, a portable computing terminal is used to divide the processing and perform data analysis for each small sensor locally, thereby reducing the transmission of data for large images and improving processing efficiency.

Recently, Raspberry Pi 4 (*Raspberry Pi*, n.d.) has become one of the most popular microcomputers in the world owing to its portability, size (of a credit card), extremely low power consumption, and complete internal structure. In addition, Raspberry Pi uses the open-source Linux system as its main operating system, which makes it extremely scalable, and it can realize the required functions with the help of many open-source projects.

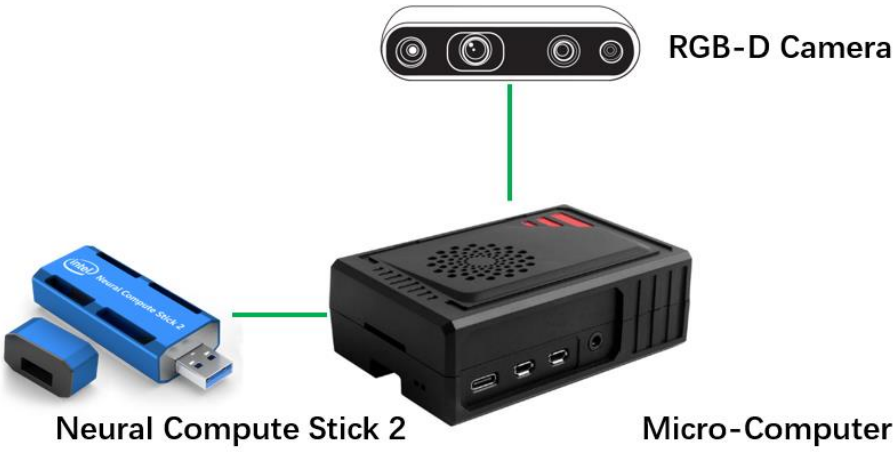


Figure 8 Portable computing terminal.

However, Raspberry Pi also has some shortcomings. For example, because of its low power consumption, it cannot drive some large GPUs for deep learning training, and the integrated graphics card that it carries cannot do this work as well. Thus, in this research, the Neural Compute Stick 2 (NCS2) produced by Intel was used as an external neural network accelerator to make up for this shortcoming. This device can accelerate neural network inference operations at relatively low operating power. NCS2 needs to be used in conjunction with Intel OpenVINO Toolkit (*Intel*, n.d.). This is an open-source software developed by Intel, which mainly includes the Model Optimizer and an inference engine to receive neural network architecture and weight information.

Another important part is the visual information collection unit. In this study, the RealSense R435i depth camera produced by Intel was used to collect 2D image information and 3D depth information.

Compared with centralized computing, the portable computing unit has better scalability and reduces the requirements for data transmission. It can perform real-time processing at the sensor and transmit the results to the central processing unit for rapid analysis. In some large construction sites, where a large number of sensors need to be deployed, the computational burden on the central processing unit can be effectively relieved. At the same time, because of the mobile computing unit's high portability, real-time processing, and early warning capabilities, it can be deployed at the dead ends of construction sites, blind areas of large operating vehicles, and other accident-prone areas, thereby speeding up rescue after accidents and reducing secondary injuries.

3.1.4 IMU-based human motion recognition

This section introduces IMU-based human motion detection, in which IMUs measure triaxial (3D) accelerations and triaxial angular velocities. This approach can also easily obtain information directly without numerous restrictions.

In this proposal, motion capture during work activities is mainly considered, so the body's four limbs are the observation focus. The limbs can express most of the essentials of movements. Despite minimizing the number of wearable devices, this research still maintains the detection accuracy of basic and relatively complex movements.

IMU motion recognition is based on a previous study (Dehzangi & Sahu, 2018), which introduced a human activity recognition method in a normal environment; the activities they considered are walking, walking upstairs, walking downstairs, sitting, standing, and sleeping. In this proposal, because the subjects were construction workers, new motions are added: lifting objects (one or two arms are elevated), picking up heavy objects (the swing amplitude of both arms is reduced and stiff), holding up heavy objects (the arms are partially angled and stiff), raising arms (arms are at right angles to body), regular cyclical movement (arms making a circular motion), bending over (leaning forward or backward), and kneeling (one or both knees).

The framework of the IMU-based human motion recognition system is as follows. First, relevant data is collected from users. Each motion is divided according to the time axis, and information such as the acceleration changes of the four sensors during the time is obtained. At the same time, for different motions, feature extraction is performed on the path changes of each sensor, and the activity label is created. Finally, by comparing the new action with the label data in the database, the action with the highest similarity is output as the result.

Each worker wears 4 sensors, and each sensor owns a separate port identification code. The combined data group from the 4 sensors will be judged as one worker. All four IMU sensors continuously record data from each worker, and the motion recognition method analyzes amplitude changes. At the same time, when a motion process changes dramatically, the differences between triaxial accelerations and angular velocities before and after the change are counted and recorded as change graphs. Finally, the differences are compared with a motion database to identify the best match. The sensors can not only collect motion data, but they can also collect information about the workers' surrounding environment, such as temperature, height, and air pressure, by exchanging data with environmental sensors to ensure that workers are in a proper working environment.

3.1.5 Multi-sensor fusion and analysis

Normally, because visual signals and IMU electronic signals are quite different, it is difficult to make a comparison between them. In this proposal, both the depth camera-based method and IMU sensor-based method can obtain results independently, but when it comes to some specific circumstances, such as self-occlusion, using only one kind of signal will cause a high error rate and affect the whole system.

In this research, two kinds of signals are cooperatively used by reducing the ad-vantages and disadvantages of each to further improve accuracy. The detection area of the depth camera is considered a huge 3D coordinate system. The depth camera is placed on one side of the system, and IMU sensors are also calibrated before loading to make sure they are consistent at time 0.

As described above, when recording starts, both the camera and IMU sides generate constant 3D coordinate changes. For the depth camera side, the variation and value of specific points are obtained from the coordinates in the depth map and frame platform. For the IMU sensor side, during movement the three axes change with different accelerations, and, using the origin set at time 0, the path changes and distance are calculated by Equation (3.1).

$$\vec{S} = \int(\int(\vec{a})dt)dt, \tag{3.1}$$

where $\vec{S}$ represents directional distance and $\vec{a}$ represent average acceleration during period t. Although the units, distance, and size are quite different between depth map coordinates and IMU sensor coordinates, it can be described that the change amplitude curve between each set of specified coordinate points (in this case, points of two elbows and two knees) by considering the weight of each kind of sensor. Thus, a more accurate result is output for comparison with the database, resulting in higher reliability for human motion recognition.

The final degree of change is shown in Equation (3.2).

$$\Delta P = \frac{\frac{\Delta P_v}{P_v^0}\cdot\alpha + \frac{\Delta P_I}{P_I^0}\cdot\beta}{2}, \tag{3.2}$$

where ΔP_v is the change of motion from the visual side, ΔP_I is the change of motion from the IMU sensor side, P_v^0 and P_I^0 are the initial states of the current time segment, and α and β are weight coefficients for visual and IMU sides, respectively.

In this study, since the effectiveness of existing sensors cannot be evaluated in advance, in this experiment, the weight distribution between different sensors is relatively ideal, and in this experiment, they are distributed into 50%, and 50%.

Figure 9 shows a flow chart of the cooperative method for fusing both data streams. The key point of this research is how to compare the data of the visual sensor and IMU sensor at the same latitude. In this experiment, a concept called “degree of change” was proposed, it is described as the change amplitude curve between each set of specified co-ordinate points (in this case, points of two elbows and two knees) by considering the weight of 2 kinds of sensors. The recognition based on the visual sensor will mark the interest points through skeleton detection, and the pixel path in 3d coordinates can be calculated. The recognition based on the IMU sensor will collect the acceleration and

angular velocity, by using the double integration method, the distance and path in 3d coordinates can be calculated. Finally, the degree of change formula is used to obtain the degree of change of each interest point, to compare with the database, and to find the best match.

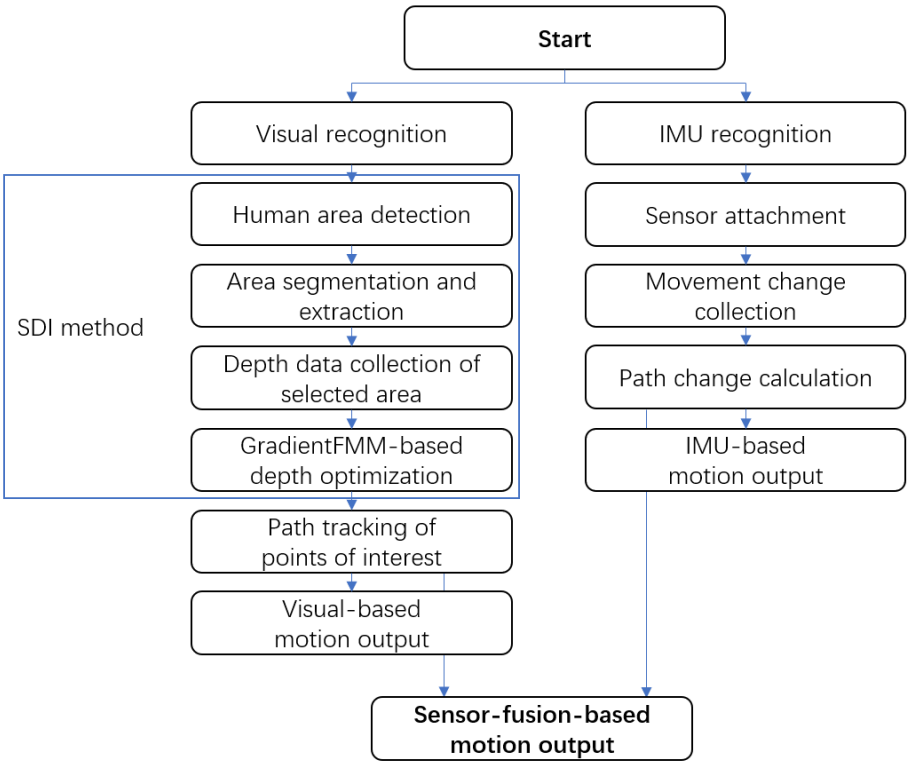


Figure 9 Flow chart of the proposed multi-sensor-based motion recognition.

3.2 Hardware and details

3.2.1 RGB-D camera

The Realsense D435i is a stereo depth camera that combines depth sensing with RGB (color) imaging, enabling accurate and real-time 3D scanning and sensing in a variety of applications. It is a product of Intel and has been designed for use in various fields such as robotics, drones, AR, VR, and autonomous vehicles.

Figure 10 Realsense D435i used in this experiment.

The D435i features a global shutter that enables accurate image capture, even in fast-moving objects. It also has an infrared projector and a stereo camera that enable depth sensing and object tracking with high accuracy. Additionally, it has an IMU (Inertial Measurement Unit) that provides precise motion-tracking data, which is useful for navigation and robotics applications.

The camera can be easily integrated with various platforms and operating systems, including Windows, Linux, ROS, and OpenCV. It also supports multiple programming languages such as C++, Python, and Java, making it a versatile solution for developers and engineers in various industries.

In this experiment, Realsense D435i shown in Figure 10 is the main solution for human action recognition.

3.2.2 Micro computer

To ensure the portability and scalability of this method, it used a microcomputer as the processing terminal for the data. In this experiment, Raspberry Pi 4 has been used.

The Raspberry Pi 4 is the latest iteration of the popular Raspberry Pi series of single-board computers. The Raspberry Pi 4 features a quad-core 64-bit ARM Cortex-A72 processor with clock speeds of up to 1.5GHz, which provides improved performance compared to its predecessors.

In terms of connectivity, the Raspberry Pi 4 has dual-band 802.11ac wireless, Bluetooth 5.0, Gigabit Ethernet, two USB 3.0 ports, two USB 2.0 ports, and two micro-HDMI

ports for connecting displays. It also has a 40-pin GPIO header, which allows users to connect various sensors, actuators, and other devices.

3.2.3 Neural compute stick 2

The Neural compute stick 2 (NCS2) is a small USB-based device designed to accelerate deep learning and neural network computations on edge devices. It is a product of Intel and is intended for use in a variety of applications, including computer vision, natural language processing, and speech recognition.

The NCS2 features the Movidius Myriad X vision processing Unit (VPU), which is capable of performing up to 4 trillion operations per second (TOPS) of deep neural network inference. This makes it an ideal solution for running deep learning models in real-time on low-power devices such as Raspberry Pi or other embedded systems.

In this experiment, NCS2 is used to accelerate motion recognition in a micro-computer.

3.2.4 IMU sensor

To capture the 3D motion data of workers, this research used two kinds of WitMotion IMU sensors to ensure the accuracy of the experiment. For the knee sensor, this research used model BWT901CL, and for the elbow sensor, model WTGAHRS2 is applied. The details of the sensors will be introduced in the next section.

3.3 Experiment

3.3.1 Setup

This section describes a simulation to validate the feasibility of the proposal. It includes the following aspects:

I. Using a depth camera along with the SDI method to identify possible humans in an area and detect the human skeleton, optimize the depth map of the human area, and collect the path change of interest points in a 3D coordinate environment.

II. Using attached IMU sensors to detect the human skeleton based on four points of interest, and to record the coordinate differences to identify different motions.

III. Using the weight-based multi-signal fusion correction approach generates coordinate differences from each frame and frequency and then outputs accurate position information.

To obtain the depth maps of subjects, a portable computing terminal was constructed from a Raspberry Pi 4 as a computing unit, an NCS2 as a training accelerator, and an Intel RealSense D435i depth camera as a data collecting unit positioned at one side of

the experiment area. This camera can achieve smooth video streaming with 848 × 480-pixel resolution at 30 frames per second. The possible depth detection ranges from 0.5 to 16 m.

To obtain the 3D motion data of subjects, four WitMotion IMU sensors were attached to both elbows and knees of each subject. At the outset of this experiment, sensors were selected with the goal of detecting worker safety from multiple angles, including environ-mental factors such as high-temperature work (heat stroke) and high-altitude work (falling, etc.). This consideration led to a different selection of sensors. Sensors for elbows are closer to the heart and head, so models that detect altitude, temperature, barometric pressure, and location information are used. For the knees, a functionally simpler model was used. The choice of the sensor did not affect the results of this experiment. The knee sensor model was BWT901CL, this sensor allows USB and Bluetooth 2.0 as its transmission method, and the detectable distance can reach 10 m at most. Its baud rate of it is 115200Bd, and the sampling rate of it is 60Hz. Meanwhile, the battery life of this sensor is about 4 hours, which is suitable for most workers. This research used the WTGAHRS2 sensor for the elbows, it can provide a more accurate 3-axis inspection than the knee sensor, the baud rate of it is 9600Bd, and the sampling rate of it is 30Hz. Yet, the elbow sensor does not have a Bluetooth module, this problem can be solved by connecting to an external Bluetooth de-vice. Currently, the diversity of collected data is balanced with its low portability, in the future, this sensor will be considered to change to a more portable kind. Each IMU will provide a three-axis acceleration detection, making four of them 12-dimensional detection on the object. These IMU sensors can detect the above parameters plus air pressure and elevation, which can help to verify that the environment surrounding workers is stable and comfortable. A picture of the IMU sensor architecture is shown in Figure 11.

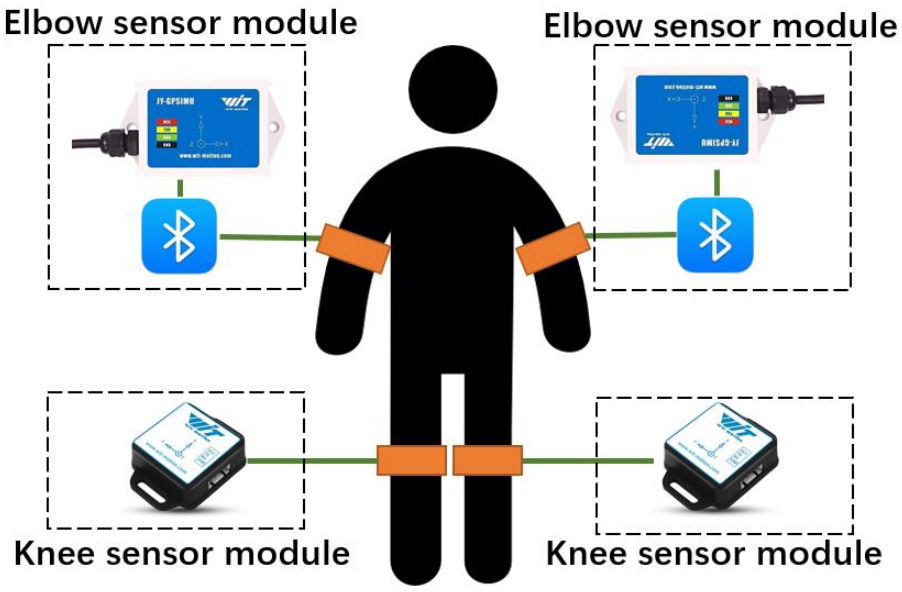


Figure 11 IMU sensor architecture.

In this experiment, the visual sensor is directly connected to the microcomputer via USB, to provide stable large-capacity image and video transmission. The IMU sensors are connected via Bluetooth, through the built-in and external Bluetooth module, to collect motion data in real-time. Each microcomputer can be considered as a local terminal, in the real-world, environment, multiple local terminals will be connected by Wireless Local Area Network (WLAN), to provide multi-point monitoring.

Before the beginning of the experiment, a database of actions is tested, which is also considered the comparison group. In this experiment, several motions were considered, including normal motions such as standing, sitting, and lying down, and motions specific to construction work, such as lifting objects, picking up heavy objects, holding up heavy objects, raising arms, regular cyclical arm movements, bending over, and kneeling. The reason to apply these motions is, in actual work, the possible motions of workers will be more complicated than in daily life, and many accidents also arise because of these actions (injuries to hands, waist, knees, etc.). By strengthening the monitoring and identification of the special motions, it is possible to effectively and quickly find out workers who continue to be in abnormal motion, thereby avoiding more serious accidents.

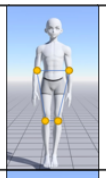
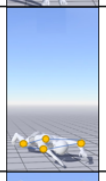
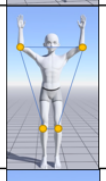
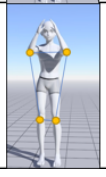
Standing		**Standing** is described as the arms drop naturally, swing slightly, the legs are stand and relaxed, and the body is not stiff.
Sitting		**Sitting** is described as sit down naturally, facing forward with a chair underneath, with legs at right angles, and arms hanging down naturally.
Lying down		**Lying down** is described as all body parts are touched or close to the floor and lie on the ground.
Lifting objects		**Lifting objects** is described as stand relaxed and raise hands naturally, with elbows approximately flush with position of ears.
Picking up heavy object		**Picking up heavy objects** is described as lifting heavy objects with both hands.

Figure 12 A detailed description of detected motions (part 1).

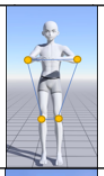
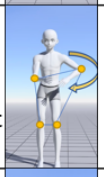
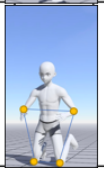
Holding up heavy object		**Holding up heavy objects** is described as embracing the heavy object with both hands, the body is leaning forward, and knees are slightly bent, and the body is stiff.
Raising arms		**Raising arms** is described as stand naturally with hands raised from both sides, flush with shoulders.
Regular cyclical movement		**Regular cyclical movement** is described as lean forward slightly, fix the object with the left hand, and make a circular motion with the right elbow at the up/down and front/back angles.
Bending over		**Bending over** is described as similar to a bowing motion, lean forward, and naturally close hands at body sides, showing a movement that is not too rigid.
Kneeling		**Kneeling** is described as kneel naturally, with thighs still nearly straight up, facing forward, and leaning forward slightly.

Figure 13 A detailed description of detected motions (part 2).

Here, this research gives a detailed description of the considered motions. Of these 10 motions, three are normal motions and seven are specific to construction. All motions are considered as starting from facing ahead. Among the normal motions, standing, sitting, and sleeping (lying down) are included, which is also commonly seen in other studies. Construction motions are motions that are commonly seen at construction sites, especially when workers are working in narrow areas or when they must reach a height that a normal standing person would find hard to reach. A mirror schematic diagram of these motions is shown in Figure 12 and Figure 13, where the yellow circles in the figures are the points of interest where sensors are attached.

3.3.2 Area layout

The data collecting and experiment area is shown in Figure 14 and Figure 15. The portable computing terminal was placed at the front of the whole area, the IMU sensors were attached to the subject's elbows and knees, and then the workers made the 10 motions (Table 1) in front of the depth camera.

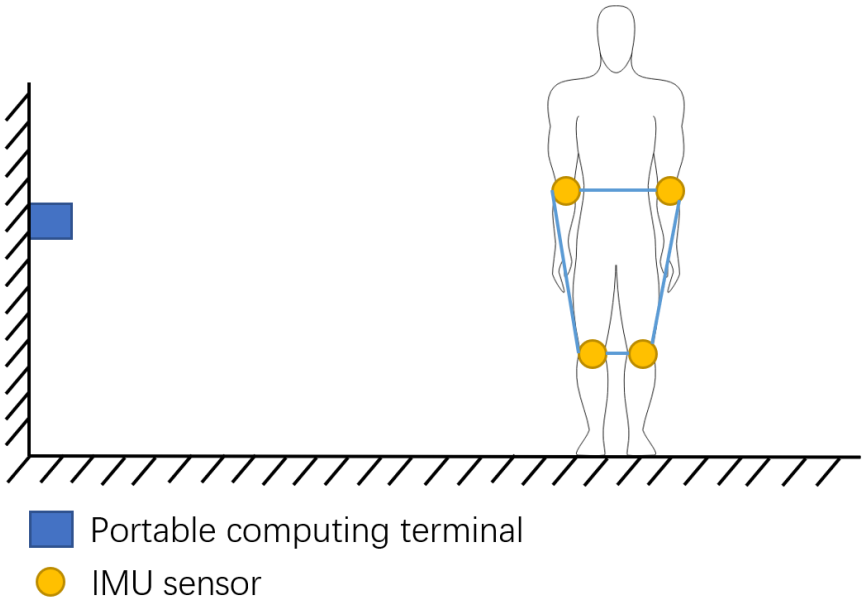


Figure 14 The sensor arrangement and environment relationship diagram. (Front view)

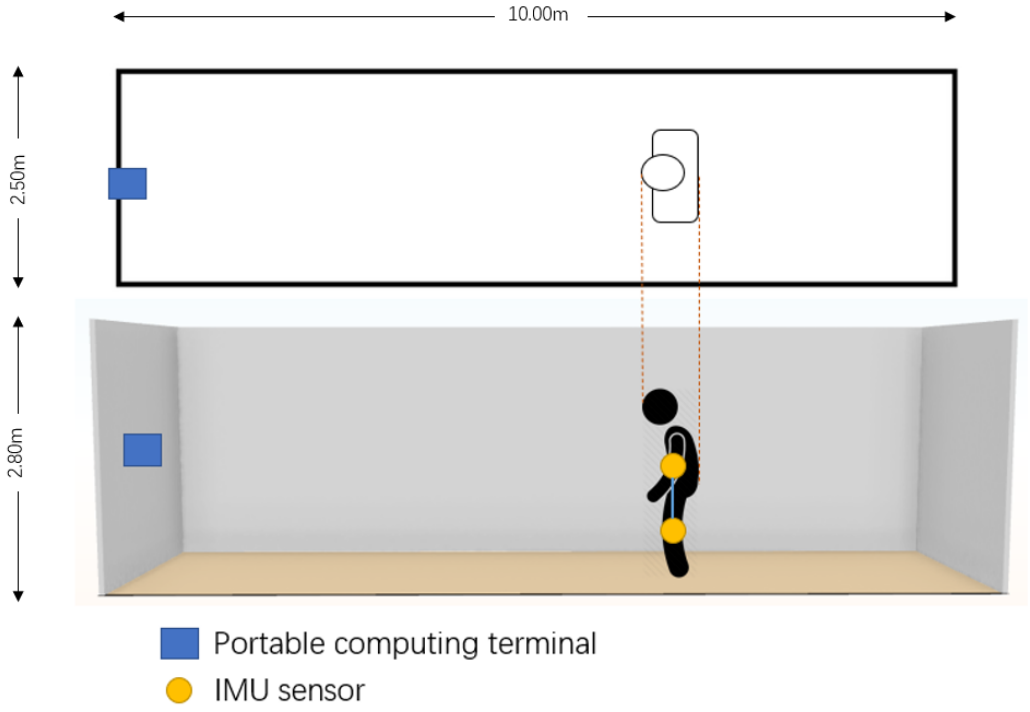


Figure 15 Experimental scenes and character relationship diagram. (top view and 3D front view)

3.3.3 Procedure

The experiment procedure is as follows:

Step 1. First, the sensors were attached to the experiment subject (worker surrogate), and the subject stood in different positions inside the experiment area to test the effectiveness of the SDI method.

Step 2. Next, the subject performed the 10 motions in order, with a short pause between every two motions. (The data collecting process is done during the construction of the dataset)

Step 3. Then, on the visual side, the possible human area was detected and the depth information inside the selected area was collected. Depth optimization based on GradientFMM filled in empty points inside the selected area and labeled points of interest. The coordinate amplitude of points of interest was also recorded.

Step 4. Next, on the IMU side, data changes from the four sensors were during the process, and a low-pass filter was employed to eliminate redundant noise.

Step 5. Then, the acceleration changes of each sensor were used to calculate path changes by the double-integral method.

Step 6. Changes in the points of interest, human elbows, and knees, from both the visual side and IMU side, were calculated separately to obtain the degree of change within a certain period.

Step 7. The degree of change from both the visual and IMU sides was used to calculate the average weighting, and the result was compared with the database to find the best match.

Step 8. Finally, the similarity from the visual side, IMU side, and sensor fusion side was compared to determine whether the sensor fusion method showed the best result.

To process the degree of change measured by the IMUs, this research set the experimental data processing environment as follows:

Step 1. Calibration procedure: The output offset component of the acceleration sensor was removed because of the presence of static acceleration (gravity). Then the acceleration was averaged when the accelerometer was not detecting motion (the collection of more samples improved the accuracy of the calibration result).

Step 2. Low-pass filtering: Signal noise in the accelerometers (both mechanical and electronic) was eliminated to decrease the error while integrating the signal.

Step 3. Mechanical filtering: When in a stationary state, small errors in acceleration were treated as constant speeds, which indicates a continuous movement and

unstable position, affecting the actual motion detection. A mechanical filtering window helped to distinguish these small errors.

Step 4. Positioning: The acceleration of each time period was known, and this research used the double-integral method to obtain distance information. The first integral gain speed and the second gain distance were applied to obtain the position.

In the simulation, the 10 motions were expected to have their unique degree of change, including vector changes from the left and right elbows and knees:

$$M_n^T = \left[\overrightarrow{A_n}, \overrightarrow{B_n}, \overrightarrow{C_n}, \overrightarrow{D_n}\right], n \in [1,10], \tag{3.3}$$

$$\overrightarrow{W_n} = (\Delta x_n^W, \Delta y_n^W, \Delta z_n^W), W = A, B, C, D, \tag{3.4}$$

where A, B, C, and D represent the left elbow, right elbow, left knee, and right knee, respectively, and $\overrightarrow{W_n}$ represents the change of motion from the four points of interest.

From the visual side, the collected point of interest data included pixel position and depth information, which output a vector change. From the IMU side, through the acceleration of three axes and time, Equation (3.3) was used to obtain the distance in all directions, thereby obtaining the vector change. Then, a weighting coefficient was assigned to the visual and IMU sides through a standard normal distribution. Next, Equation (3.4) is used to calculate the integrated vector change:

$$F^T = \left[\frac{\frac{A_v}{A_v^0}\cdot\alpha+\frac{A_I}{A_I^0}\cdot\beta}{2}, \frac{\frac{B_v}{B_v^0}\cdot\alpha+\frac{B_I}{B_I^0}\cdot\beta}{2}, \frac{\frac{C_v}{C_v^0}\cdot\alpha+\frac{C_I}{C_I^0}\cdot\beta}{2}, \frac{\frac{D_v}{D_v^0}\cdot\alpha+\frac{D_I}{D_I^0}\cdot\beta}{2}\right], \tag{3.5}$$

where A_v is the left elbow vector change on the visual side and A_I is the left elbow vector change on the IMU side. F is compared with M_1 to M_{10} in Equation (3.5) to find the highest similarity.

Regarding the construction of the dataset, this research recorded motion data from 5 male adults between the age of 20 and 30. For each motion, 50 pairs of coherent and clearly behaved data from the perspective of visual and IMU sensors are collected. This process is done twice because the data from two different distance intervals are required. From the visual aspect, each motion is labeled based on the features such as the coordinate change of points of interest and depth information, from the IMU aspect, each motion is labeled based on the path change of each sensor attached.

This study used the same simulation method that was used in one of the previous paper (T. Chen et al., 2020). For each motion, 100 pairs of sample data from each distance interval were prepared. The sample data were generated based on the dataset by adding

random interferences and white noises, to simulate deviations caused by the effects of real data collection. Weight coefficients obeyed a standard normal distribution.

Table 1 Description of simulation situations.

	Detecting distance	Application of the SDI method
Situation 1	4-6m	×
Situation 2	8-10m	×
Situation 3	8-10m	√

As shown in Table 3.1, three situations for human motion detection were considered in this simulation. To determine whether the SDI method applied to human motion detection can succeed at a relatively long distance, the results for these three situations are compared and discussed.

3.4 Result

By combining the visual pixel changes based on the depth camera and acceleration path changes based on IMU sensors, this research generated graphs with the highest similarity of each set of sample data. Some typical motion captures and their results of similarity are shown in Tables 2, 3, and 4, and Figures 16, 17, and 18.

Figure 3.10 shows the result obtained during situation 1, in which the subject stood at a highly detectable distance (around 4-6 m away from the camera) and the SDI method was not applied. The formulas for recall and precision are shown in Formula (3.6) and (3.7).

$$\text{Recall: } R_{ec} = \frac{TP}{TP+FP}, \tag{3.6}$$

$$\text{Precision: } P_{re} = \frac{TP}{TP+FN}, \tag{3.7}$$

A recall is for the original sample, which indicates how many positive examples in the sample are predicted correctly. There are also two possibilities, one is to predict the original positive class as a positive class (TP), and the other is to predict the original positive class as a negative class (FN). Precision is for the prediction results, and it indicates how many of the predicted positive samples are true positive samples. Then there are two possibilities for predicting positive, one is to predict the positive class as the positive class (TP), and the other is to predict the negative class as the positive class (FP).

Table 2 Simulation result for situation 1.

Real motion / Detected motion	standing	sitting	lying down	lifting objects	picking up heavy	holding up heavy	raising arms	regular cyclical movement	bending over	kneeling	Precision
standing	93	1		1	2	1		1			93.94%
sitting		97									100.00%
lying down			98							3	97.03%
lifting objects				87	5	7	2				86.14%
picking up heavy	4			4	87	3					88.78%
holding up heavy	3	2		8	5	89	5				79.46%
raising arms					1		93				98.94%
regular cyclical movement								95	3		96.94%
bending over								4	97	1	95.10%
kneeling			2							96	97.96%
Recall	93%	97%	98%	87%	87%	89%	93%	95%	97%	96%	93.27%

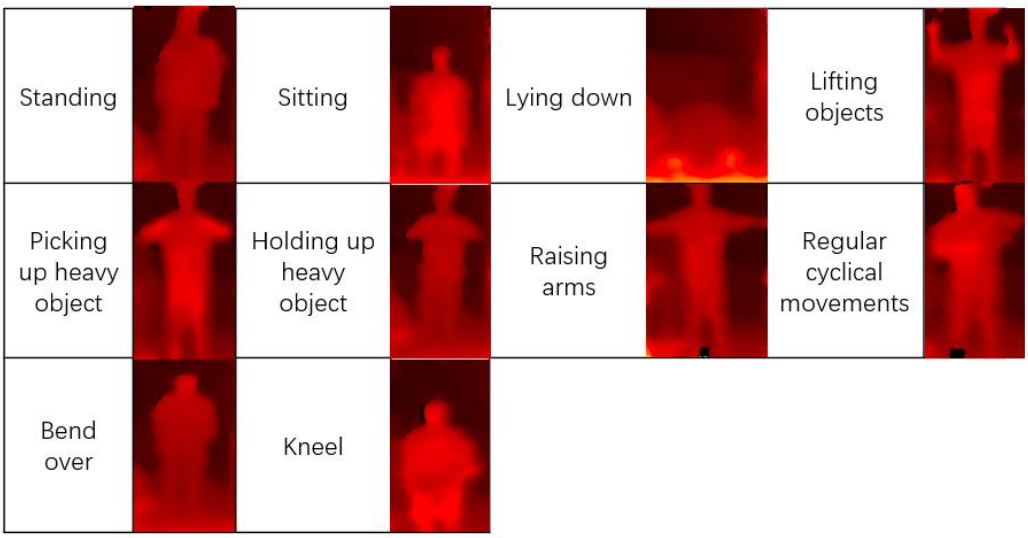


Figure 16 Example group of optimized depth motion capture for situation 1.

It can be learnt from Table 2 that most of the motions can be correctly identified, but in some motions, some points of interest have similar paths, which can be confused with other motions, resulting in wrong outputs. Figure 16 shows the example group of optimized depth motion captures for situation 1.

Table 3 shows the result obtained during situation 2 when the subject stood at a relatively long distance (around 8-10 m away from the camera) and the SDI method was not applied. Figure 17 showed that when people stand a long distance away from the camera, the depth detection does not perform well, especially when the detection requires the identification of similar motions with different depth changes, such as bending over, regular cyclical movement, lifting, and picking up or holding heavy things. The reason is, the further the object is from the depth camera, the higher the RMS error will be, even after optimization, some of the motions still cannot be identified because of errors and interference. If the fusion results are not corrected by the IMU sensor results, the accuracy could be lower.

Table 4 shows the result obtained during situation 3 when the subject stood at a relatively long distance (around 8-10 m away from the camera) but the SDI method was applied. It can be learnt from Figure 18 that some depth-based errors are well corrected by the SDI method using RGB person scale finding, and skeleton assistance based on PoseNet, which provide good separation for bending over and kneeling, and improve the identification between lifting and holding up or picking up heavy objects. There are still some problems while using the SDI method due to the long distance, and the skeleton or object scale is sometimes poorly constructed. In addition, the conflict between and misjudgment of some points of interest may lead the result to a completely unrelated motion, such as bending over, which was misidentified as five different motions several times.

Table 3 Simulation result for situation 2.

Real motion / Detected motion	standing	sitting	lying down	lifting objects	picking up heavy	holding up heavy	raising arms	regular cyclical movement	bending over	kneeling	Precision
standing	90	2		2	2	3	1	1	3		86.54%
sitting		84				2			7	4	86.60%
lying down			95								100.00%
lifting objects	3			75	10	7	8				72.82%
picking up heavy	4			15	65	19					63.11%
holding up heavy	3	6		8	22	68	5				60.71%
raising arms					1		86				98.85%
regular cyclical movement						1		82	13		85.42%
bending over		8	2					17	77	11	66.96%
kneeling			3							85	96.59%
Recall	90%	84%	95%	75%	65%	68%	86%	82%	77%	85%	81.07%

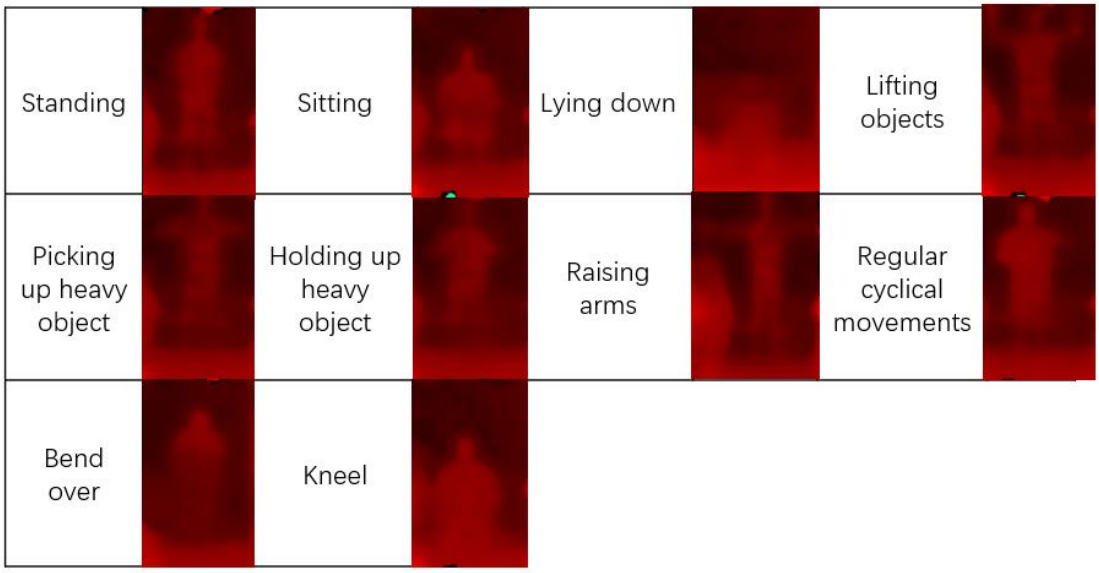


Figure 17 Example group of optimized depth motion capture for situation 2.

As for the validation of the result, this research considered the measure of precision and recall. Precision is based on prediction results, it indicates how many of the "predicted positive instances" are truly positive, recall is based on original samples, and it indicates how many positive instances are predicted correctly. As the final criterion, this research adopted the F1-measure approach, which is the weighted harmonic average of precision and recall. In Table 2, the recognition result shows that the precision and recall are 93.43% and 93%, and the F1-measure is 93.27%, the similar results suggest a good classification outcome. The evaluation standard in Table 3 and Table 4 is the same as in Table 2. In Table 3, the precision, recall, and F1-measure are 81.76%, 81%, and 81.07%; in Table 4, the precision, recall, and F1-measure are 92.86%, 93%, and 92.80%, also support the good classification outcome.

A comparison of the different detection situations is shown in Figure 19, from which can the result be seem, when the subject is far from the camera without the assistance of the SDI method, the visual side detection rate is much lower than that of the other situations. Because the Bluetooth transmission quality of the IMU sensor is also affected at long distances, the final accuracy is not very high. With the SDI method applied, long-distance motion detection can reach the accuracy of short-distance detection, and the lack of depth information is effectively compensated by image processing.

Table 4 Simulation result for situation 3.

Real motion / Detected motion	standing	sitting	lying down	lifting objects	picking up heavy	holding up heavy	raising arms	regular cyclical movement	bending over	kneeling	Precision
standing	98			2	2	4			2		90.74%
sitting		97						3			97.00%
lying down			98							3	97.03%
lifting objects				91	4	4	2	2			88.35%
picking up heavy	1			5	89	3			2		89.00%
holding up heavy	1	1		2	3	89	3		2		88.12%
raising arms					2		95		3		95.00%
regular cyclical movement								88	4		95.65%
bending over		2	1					7	87	1	88.78%
kneeling			1							96	98.97%
Recall	98%	97%	98%	91%	89%	89%	95%	88%	87%	96%	92.80%

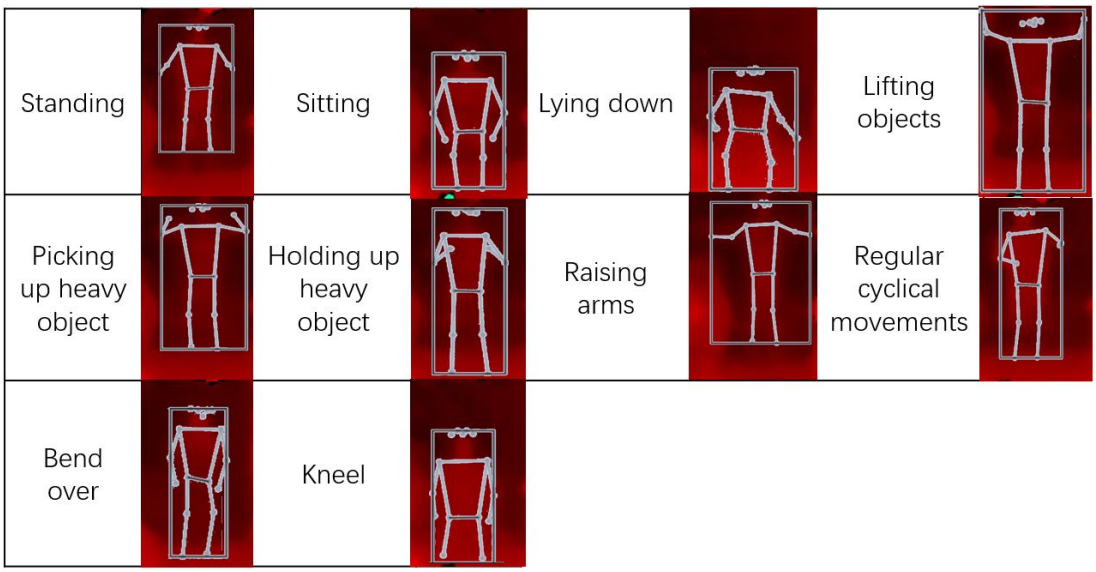


Figure 18 Example group of optimized depth motion capture for situation 3.

Our experiment shows that the average accuracy of motion recognition by multi-sensor fusion at a short distance (situation 1) and relatively long distance assisted by optimization from the SDI method (situation 3) can reach 93.27% and 92.80%, compared with situation 2, the accuracy improved about 12%. Although the number of samples is not large, this result shows that the proposed SDI method based on multi-sensor fusion makes it possible to realize high-precision motion recognition beyond the optimal recognition distance of the depth camera, and it can detect the different kinds of motions used at construction sites.

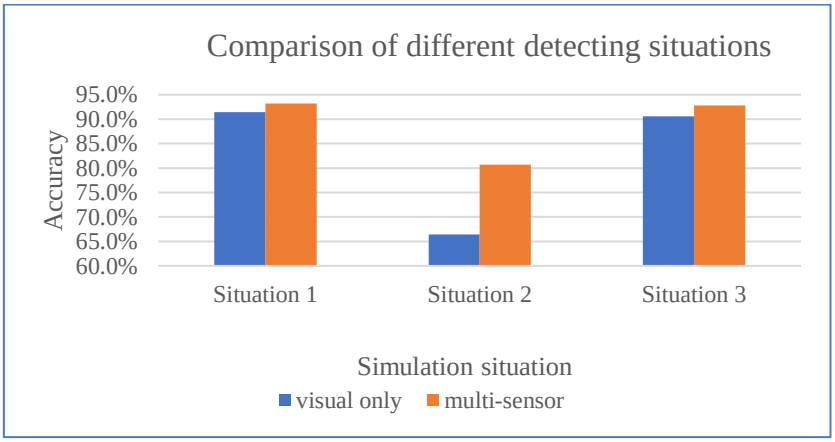


Figure 19 Comparison of different detection situations.

Currently, the assumed applicable environment of the proposed method can be described as some medium-sized indoor areas of the construction site or buildings, some corners or blind spots of the construction site. Due to environmental limitations, the workers that work in a blind spot, a closed or semi-closed area can be easily ignored, yet the danger occurrence rate of these places is very high, and because of the remote location, the rescue is not timely as well. Due to the different prerequisites, the proposed

method is different from the mainstream construction site monitoring approach, and the available situations are relatively limited, but the importance is still not to be ignored.

3.5 Summary

This experiment proposes a novel method for motion recognition and hazard prevention of construction workers using an integrated sensor network. The objective is to achieve time warning and rescue in dead ends and blind areas of different construction situations by monitoring workers. To effectively ensure the safety of workers in the complex environment of a construction site, this research extended the detection distance of a normal depth camera. This research improved the depth camera-based motion recognition with SDI, which uses preprocessing on human scale finding and depth map optimization methods to effectively reduce the detection errors and calculation burden of a broad range of depth data, while at the same time enhancing the recognition distance and accuracy of the depth camera in a selected area. A portable computing terminal is also used instead of a single depth camera to achieve local analysis, avoiding the computing burden caused by transferring a large amount of data to a central processing unit. This research also demonstrated that using different types of sensors to recognize human motion improves the accuracy of motion recognition.

The proposed methodology has some limitations. To simplify the problem, this research limited the considered motions in the motion recognition method to detect only 10 selected types of motion. Also, owing to the hardware limitations of the microcomputer, the current configuration of the SDI method cannot achieve real-time detection. The experimental data and results of the simulation were collected over time but analyzed at one time. In practical applications, data collection and analysis will be processed in a short period, so that even if workers experience abnormal conditions, they can be quickly discovered and rescued. At present, the result of this approach only represents its performance in a simulated environment. Although the function is basically realized, it is not yet mature enough to be applied. The next step will be focusing on real-time realization, improving the efficiency of recognition, and the application in the real construction site environment. The results collected from an actual construction site environment are expected to have lower accuracy due to interferences.

During the simulation process, this research discovered several problems, such as the points of interest for some actions having similar trajectories, resulting in some misjudgments. Therefore, finding out how to use other methods to determine and classify similar actions more accurately to improve the performance of motion recognition will be the focus of future work. Future work will also include adding a real-time warning based on motion recognition to the detection system to realize the original intention of this method, specifically, improved construction hazard prevention.

Chapter 4

Active early warning system for heavy vehicle drivers using mixed reality

The use of heavy machinery has always been a challenging task that demands the utmost attention and concentration from operators. The increasing demand for automation in the industrial sector has led to the development of advanced technologies to ensure the safety of workers and minimize accidents in the workplace. MR is one such technology that has shown immense potential in enhancing safety and efficiency in various industries. This section proposes a novel method that utilizes an MR device to monitor heavy machinery operators in real-time to ensure their safety while driving.

The proposed method involves the use of an MR device that can obtain the operator's gaze direction, head direction, and hand condition while driving heavy machinery. The device is equipped with sensors that can detect the operator's eye movement and head orientation and can detect whether the operator is relaxed or tense. Based on these parameters, the device sets dangerous and safe zones in the user interface (UI). If the operator's gaze or head movement falls into the dangerous zone, the device immediately alerts the operator with visual and auditory cues.

The proposed method can help prevent accidents in the workplace caused by operator fatigue or distraction. It can detect whether the operator is too tired to move their head or gaze to dangerous areas, preventing accidents caused by inattention. The MR device can provide real-time feedback to the operator and warn them immediately of any potential dangers. The device's alert system can help prevent accidents caused by the operator's inattention, making the work environment safer and more efficient.

4.1 Methodology

After synthesizing all relevant studies, it can be known that most of the driver-based safety monitoring and early warning method is partial biological index measurement, which has high requirements for daily use and re-implementation. At the same time,

due to the complexity of the processing equipment, it is possible to implement the actual environment. The research on safety monitoring of the construction environment based on MR technology is based on macroscopic, and they focus more on training, danger prevention, and remote human monitoring.

In this study, a mixed-reality-based active monitoring system for drivers of heavy vehicles will be proposed, instead of concentrating on the entire environment of the construction site and biological methods, this method focuses on the dangerous behaviors that are not subject to the drivers' will, such as irregular driving due to distraction or fatigue, etc.

4.1.1 Data collection and preparation

This study is based on MR technology, using MR equipment worn by heavy vehicle drivers, to analyze the data and achieve real-time warnings.

One of the current mainstream MR headsets, HoloLens 2 (*HoloLens*, n.d.) will serve as the hardware basis for this study. Because functions such as hand motion tracking, head gyroscope, and gaze tracking are available, active driver behavior monitoring can be achieved, and multi-dimensional prediction of some abnormal and dangerous behaviors of drivers can be detected. During the monitoring process, for suspected dangerous driving behaviors, the MR device will give visual warnings (levels that won't affect normal driving), at the same time, it will also provide auxiliary reminders in auditory and other aspects. This research is mainly aimed at the slow-moving heavy vehicles in the construction area. Within the scope permitted by law, the system can be extended to roads and other areas.

4.1.2 Method details

As depicted in Figure 20, the various potential hazards that heavy vehicle drivers may encounter, such as fatigue and distraction, pose a serious threat to their safety and those around them. These risks can be identified by closely monitoring their hand gestures, head gaze, and eye gaze. Consequently, it is imperative to pay utmost attention to these abnormal behaviors in order to mitigate any potential accidents and ensure the safety of all road users. The monitoring focus can thus be divided into the following critical areas:

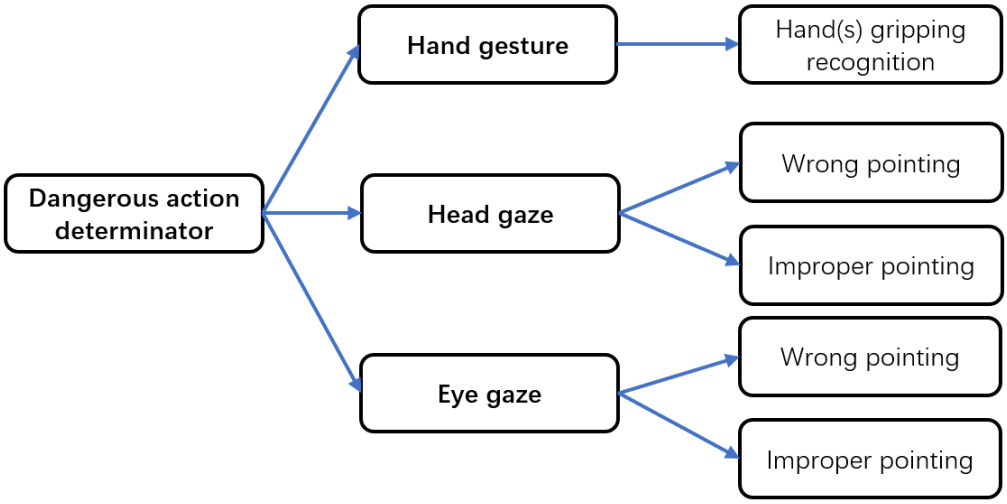


Figure 20 Introduction to dangerous actions based on three dimensions.

4.1.2.1 Hand monitoring

Based on the detailed functionality description of the Mixed Reality Toolkit (MRTK) (*MRTK*, n.d.), it is evident that HoloLens 2 possesses the ability to accurately recognize up to 25 nodes of a single hand (as illustrated in Figure 21). The hand is considered one of the primary input methods, and it offers a wide range of gesture-based interactions. These gestures include actions such as grabbing, clicking, air tapping, and long-distance cursor pointing, which can be readily executed on HoloLens 2.

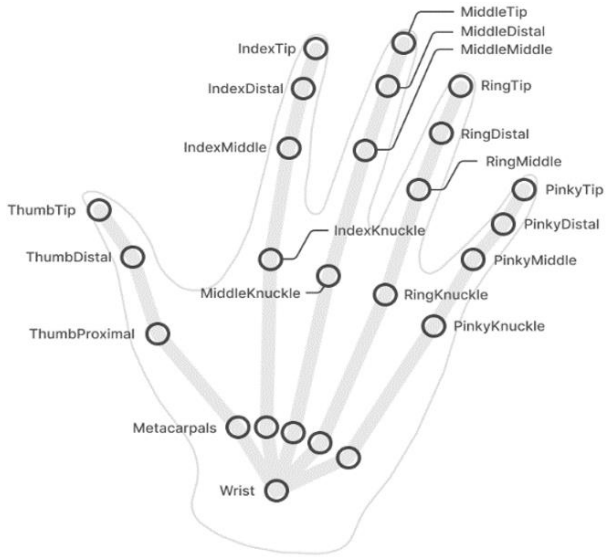


Figure 21 25 nodes that HoloLens 2 can recognize in one hand.

Our research has identified that when a driver is suspected of being distracted or fatigued, their hand movements can often become erratic, such as leaving the steering wheel. This may occur when they attempt to operate the center control panel or when their hands fail to maintain the proper position on the steering wheel. In view of the capabilities of MRTK, this research has opted to employ the grasping gesture to simulate the driver's hold on the steering wheel. By analyzing the grasping state of the driver's hands on the wheel, this research can accurately assess their current state in one dimension, as depicted in Figure 22. This approach will enable us to detect any deviations in the driver's behavior and take appropriate corrective action to ensure their safety and that of other road users.

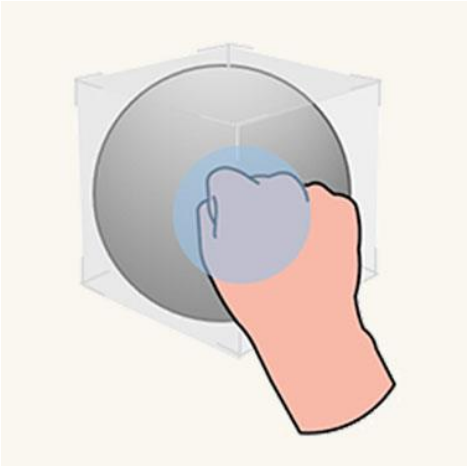


Figure 22 Description of grab motion in MR environment.

4.1.2.2 Head gaze monitoring

Given that the hands of the driver may be occupied during driving, it can become challenging to perform certain gestures. To address this issue, MRTK offers the added functionality of head gaze and stay, allowing the user to perform various operations by pointing their head toward the intended target.

In this research, the primary objective is to identify and mitigate dangerous driving behavior among drivers. Specifically, this research aims to detect instances of driver distraction and fatigue, which are two major contributors to road accidents. When a driver is distracted for prolonged periods, their head tends to point towards non-frontal areas, such as the roadside, rearview mirror, or center control panel.

To address these challenges, this research leverages the head gaze and stay functionality of MRTK to analyze the driver's head-pointing behavior and provide early warnings in cases of suspected distraction or fatigue. This proactive approach allows us to intervene before accidents occur, preventing potential injuries or loss of life. By utilizing MRTK, this research can contribute to enhancing road safety and reducing the risk of accidents caused by driver distraction or fatigue.

4.1.2.3 Eye gaze monitoring

As shown in Figure 23, the HoloLens 2 introduces a novel feature of eye-tracking, which eliminates the need for head movements while selecting various objects within the same area of view. This feature, combined with the gesture functionalities, allows for convenient object selection, movement, and manipulation.

Figure 23 Describe the range of eye recognition that hololens2 can perform.

In this research, the main purpose is to detect dangerous driving behaviors among drivers, specifically those associated with distraction and fatigue. This research has observed that these states can significantly impact a driver's eye movements. In normal driving conditions, the driver's eyes are primarily fixated on the road ahead, with occasional drifts toward the rearview mirror or vehicle instrument information. However, distracted driving can cause a driver's gaze to focus on abnormal directions or objects, leading to unintended consequences such as fine-tuning the steering wheel and driving off the road or into oncoming traffic. Similarly, when fatigued, the driver's eyes may not concentrate, leading to increased blinking due to heavy eyelids and slower refocusing times after each blink. These factors can cause delayed analysis of road conditions and result in severe accidents.

To address these issues, this research leverages the eye-tracking function of MRTK to monitor the driver's eye movements in real-time. By setting up a sensing function in the abnormal gaze areas, this research can provide visual and auditory warnings to drivers who gaze too much, thereby enabling timely alerts and sufficient preparation time for subsequent measures, such as slowing down the vehicle. This proactive approach can

help prevent accidents caused by distraction or fatigue and ultimately enhance road safety.

4.2 Experimental setup

For the purpose of conducting this experiment, the hardware used was the HoloLens 2, a mixed-reality device developed by Microsoft. On the software side, the experimental setup included Unity 3D version 2019.4.21f1 and MRTK version 2.6.0, which were utilized to create and develop the experimental environment. The scene construction involved utilizing Euro Truck Simulator 2 (*ETS 2*, n.d.) as the primary tool to evaluate the effectiveness of the implemented functions. Additional modifications and extensions were made to enhance the academic rigor of the experiment.

4.2.1 Mixed reality device

As illustrated on the left side of Figure 24, the HoloLens 2 is a second-generation mixed-reality device developed by Microsoft and released in 2019. This device boasts significant improvements from its predecessor, including a wider viewing range that is twice as large, a holographic resolution of up to 2K that provides a clearer projection of virtual objects, and greater visibility of internal details.

Figure 24 HoloLens 2 used in experiment 2.

Regarding its hardware, the HoloLens 2 is equipped with four visible light cameras for head tracking, two infrared (IR) cameras for eye tracking, a 1 MegaPixel (MP) time-of-flight depth sensor to detect depth information, and an initial measurement unit (IMU). Additionally, the device is equipped with a camera that can capture 8-MP photos and record 1080p 30fps videos.

The HoloLens 2 also supports advanced human-computer interaction features, including full-joint model recognition of both hands, enabling direct interaction with

virtual objects using bare hands. Moreover, the device supports real-time eye tracking, providing more natural and intuitive ways of interacting with virtual objects.

4.2.2 Unity 3d and MRTK

The Unity 3D version 2019.4.21f1 was used in this simulation experiment in conjunction with the Microsoft MRTK version 2.6.0 for developing MR applications. MRTK for Unity is a comprehensive, open-source development kit that facilitates the creation of spatially-aware applications that can be deployed across a wide range of platforms.

MRTK provides a variety of essential components and building blocks for MR development, including a cross-platform input system that enables developers to design natural and intuitive user interfaces that can be used across various devices. This input system supports a range of input modalities such as voice, hand and eye tracking, and spatial mapping, making it possible to create interactive and immersive user experiences. Figure 25 shows some example functions that MRTK can achieve.

Additionally, MRTK offers basic components for spatial interaction, including camera modules, coordinate systems, head gaze, text, and spatial sound effects. These features enable developers to design realistic and engaging environments that respond to user input and enable users to interact with digital objects in a natural and intuitive way.

Figure 25 MRTK Examples holographic view.

MRTK also provides support for advanced features such as eye tracking, which can be used to create more natural and immersive user experiences, and spatial mapping, which enables the creation of virtual representations of real-world environments.

In conclusion, MRTK for Unity is a powerful and flexible development kit that provides essential building blocks for mixed-reality applications. Its support for a wide range of input modalities and its cross-platform compatibility makes it an ideal choice for developers looking to create engaging and immersive user experiences in mixed-reality environments.

4.3 Simulation

The simulation described in this experiment was originally developed using Unity 3D and was subsequently installed on a HoloLens 2 device to facilitate actual MR experience feedback. The virtual scene was created in Unity to simulate the view of a truck driver, including position and distance information about the surrounding environment. The overall framework of an early warning system was also constructed as part of the simulation.

In this experiment, a visual model of a right-hand drive truck was used to represent the driver's view (shown in Figure 26). The driving visual range was divided into four parts based on driving behavior habits and the distribution of various parts in the vehicle: gaze danger zone, gaze attention zone, gaze safe zone, and hand operation zone. When the driver displays abnormal behavior, such as looking at the wrong zone for an extended period or loosening their grip on the steering wheel, they are assumed to be distracted or fatigued, and corresponding alarm measures will be activated.

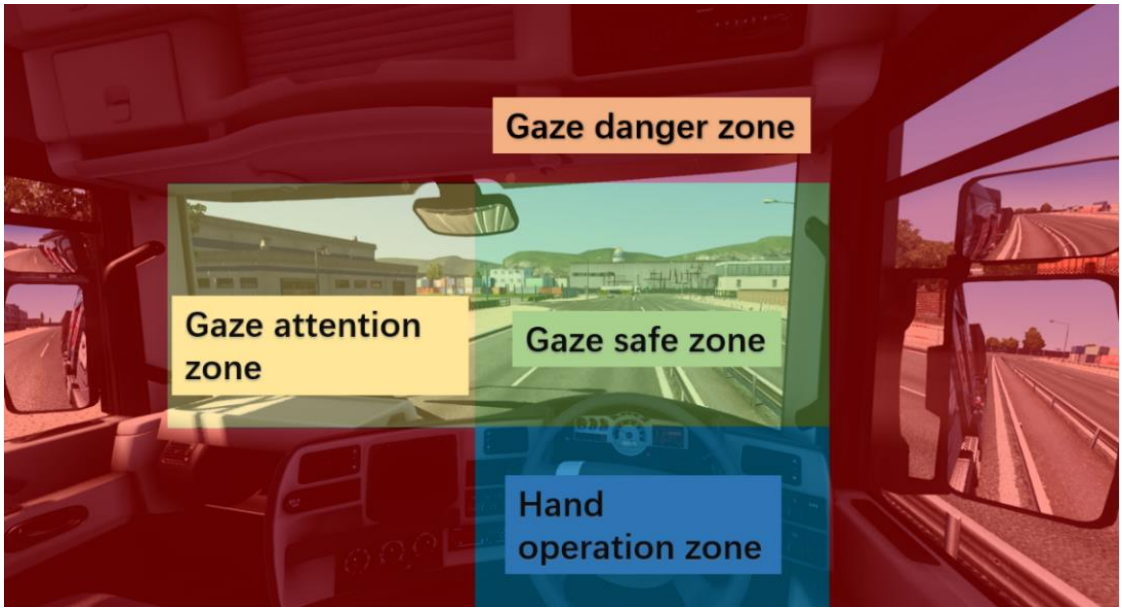


Figure 26 Introduction of division on driver’s visual area.

The simulation provides a realistic representation of a truck driver's view and behavior, enabling researchers to evaluate the effectiveness of the early warning system in detecting driver distraction and fatigue. By using a mixed-reality device, the simulation provides an immersive and interactive experience, enabling researchers to gather more accurate and reliable feedback from participants.

In general, this experiment aims to improve the early warning system for drivers in dangerous situations by adding feedback from two dimensions hand monitoring and head/eye gaze monitoring.

For hand monitoring, components not limited to *NearInteractionGrabbable* and *ManipulationHandler* are utilized to provide feedback on the driver's grasping action on the steering wheel. The sensing objects are placed about 0.35 to 0.4 meters away from the driver to ensure a natural grasping position. A new script named *HandDetector* is also added, which can detect the three-dimensional coordinates of both wrists in real-time to assist in early warning.

For head/eye gaze monitoring, components such as *EyeTrackingTarget*, *GazeProvider*, and Gaze are used to divide the different levels of areas by different colors, except for the safe zone and hand operation zone. When the driver's eyes or head is pointed to the designated area for more than 0.8 seconds, the pointed area will turn red or yellow to provide a visual alert. A short "Beep" will also provide an auditory alert. In the gaze danger zone, the "Beep" comes along with the color change to alert the driver in different dimensions as soon as possible. In the gaze attention zone, the "Beep" will come slightly later than the visual alert since the danger level is not as high as in the danger zone.

Overall, this enhanced early warning system provides drivers with additional feedback on their hand-grasping action and head/eye gaze, enabling them to respond quickly and avoid accidents.

4.4 Result

During the HoloLens 2 real machine test, the eye direction point was not shown to prevent interference with the driver. Therefore, the results of this experiment are presented in two forms. Head and hand monitoring are fed back through frames captured during the real machine test, while eye gaze monitoring is performed using Unity 3D's simulation result.

The head monitoring result is shown in Figure 27. In a simulated truck cockpit environment, the experimenter's head is pointed at each area to confirm the function by getting feedback from alerts. The tester aims his head at different zones, such as the left

and top rearview mirror, top cockpit area, center control panel, right rearview mirror, and front windshield of the co-pilot side, to receive gaze warnings and feedback.

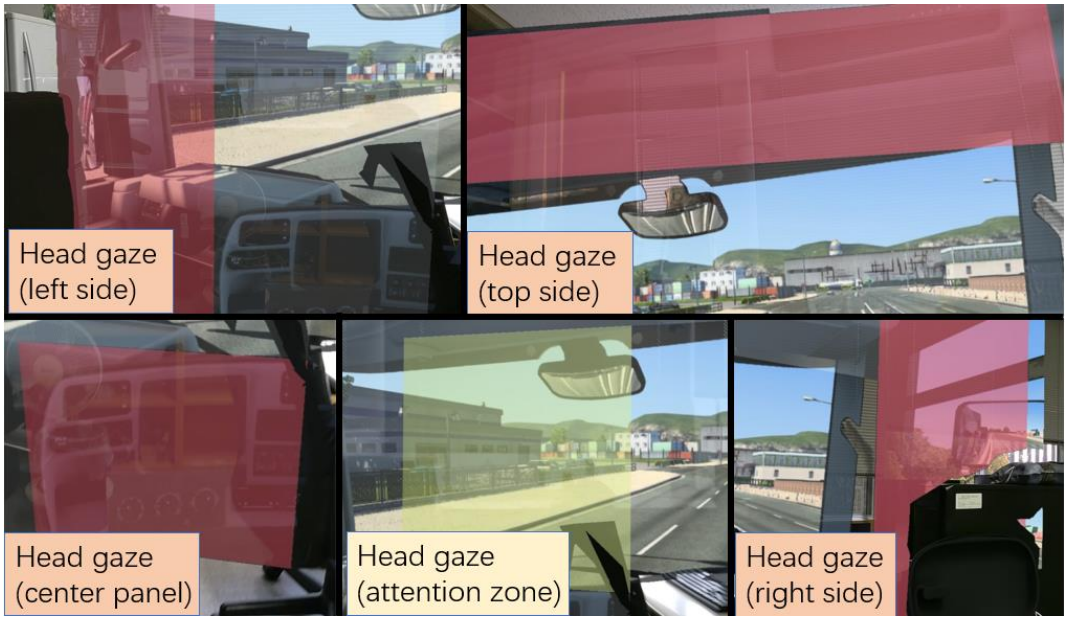


Figure 27 Alert feedback of head gazing areas.

The second part is hand monitoring, as shown in Figure 28. In a simulated truck cockpit environment, the tester's hands are used to grab the simulated steering wheel area. Two sensing objects, currently shown as red balls, are placed in the hand operation zone. When the tester grabs the right position, the sensing objects will provide feedback with a transparent effect. If the hand or both hands cancel the grasping action, the sensing objects will immediately provide an alert through color changes and auditory feedback to remind the driver to restore the correct driving posture. The position of the head-sensing object is not fixed, but it is limited inside the hand operation zone. It also has a reset function, which allows the tester to reset the object to the original position if the object's position drifts too much.

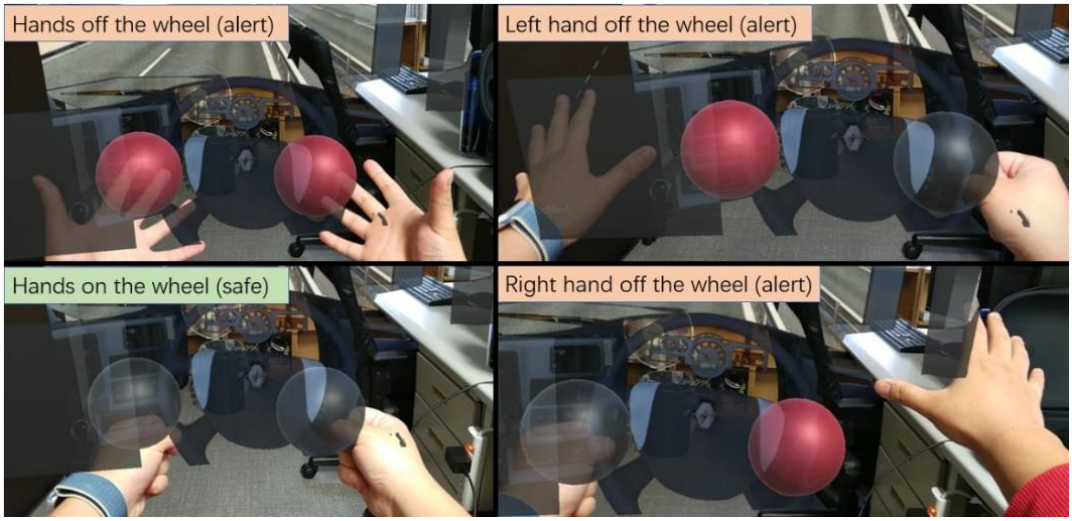


Figure 28 Grabbing motion sensing of the hand/hands in hand operation zone.

The last part is eye monitoring, as shown in Figure 29. In a simulated truck cockpit environment, the tester's head movement is fixed, and only the tester's eyes can look around the environment, stare and stay for a while in different zones. During the process of looking around, feedback from different zones is used to confirm whether the eye gaze warning function is available. The white points, marked with blue circles in Figure 9, indicate where the eyes are pointed.

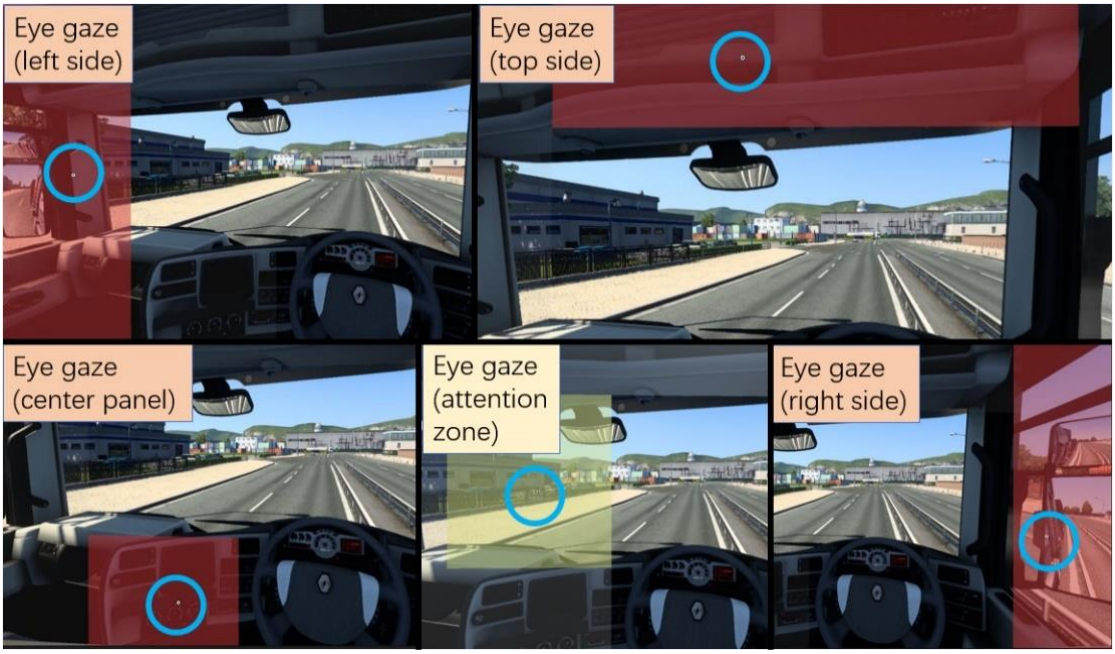


Figure 29 Alert feedback of eye-gazing areas.

4.5 Summary

This experiment proposes an MR-based active monitoring system for heavy vehicle drivers, with a focus on active early warning to detect potential distractions or excessive fatigue in real-time using HoloLens 2. The study concentrates on three frequently used body parts while driving: head, eyes, and hands, to further determine the driver's current state. Abnormal states based on these body parts are discussed, and a combination of components in MRTK and original components is used to collect and analyze information from hands, head orientation, and eye-pointing direction, providing multi-dimensional feedback to monitor abnormal movements and states of designated body parts in real-time, thereby discovering and preventing potential dangers.

The results of the experiment demonstrate the feasibility of the proposed active heavy vehicle driver monitoring system. The real-time hazard warning feedback function works well, and no unrecognizable state was observed during the tests. However, feedback from the testers highlighted the need for improvement in the UI interface design and adding richer feedback such as voice prompts. Due to hardware limitations, the realization of user-friendly products needs further improvement. Moreover, the study emphasizes the need to consider the ethics and safety concerns related to carrying MR equipment during high-concentration operations and the reliability and safety of the actual use of MR equipment, considering the various laws and regulations in different regions.

Overall, this study provides a promising solution for detecting potential hazards in heavy vehicle driving, ensuring driver safety, and preventing potential accidents. However, further research is necessary to enhance the user interface and develop additional safety measures for the practical implementation of the system. Additionally, ethical considerations must be addressed to ensure the safety and reliability of the system in compliance with relevant laws and regulations.

Chapter 5

Mixed reality-based active hazard prevention system for heavy machinery operators

The construction industry relies heavily on heavy machinery to improve work efficiency. However, improper operation of this equipment can lead to serious accidents that may result in fatalities. While many studies have developed auxiliary warning methods for blind spots of heavy machinery, these methods may not be sufficient to prevent accidents that occur when the driver is unaware or accidentally causes them. Therefore, more advanced safety measures are urgently needed to improve the overall safety of heavy machinery operations.

Recently, research based on MR technology has become popular in the construction industry. The safety management methods currently used in the industry include pre-training of on-site staff, overall monitoring based on the construction site, and pre-design of the technology of the building itself. However, this study proposes an actively monitored real-time system using MR technology that focuses on the driver's state to identify and alert them of fatigue or distracted driving to prevent serious accidents.

This study utilizes MR technology to overlay virtual UI onto the real world for more effective monitoring of the driver's state while minimizing the impact on real-world operation efficiency. The system continuously monitors the driver's behavior, including head movement, eye gaze, and facial expression, to detect signs of fatigue or distraction.

The proposed system provides significant benefits to the construction industry by improving the safety of heavy machinery operations. By continuously monitoring the driver's state and providing real-time alerts, the system can prevent serious accidents caused by fatigue or distraction. Furthermore, the use of MR technology enhances the overall efficiency of the monitoring process, improving the accuracy and speed of driver state assessment.

The proposed actively monitored real-time system using MR technology represents an innovative approach to improving the safety of heavy machinery operations. By utilizing a combination of wearable devices and in-car sensors, the system can detect signs of driver fatigue or distraction and provide real-time alerts and adaptive training modules. The system has the potential to significantly reduce the occurrence of serious accidents in the construction industry, improve worker safety, and increase overall work efficiency. The main contributions of this research are:

- The proposed method details the development of a proactive real-time monitoring system, specifically designed to prevent accidents involving heavy machinery operators;
- The proposed method is not dependent on the subjective behavior of the operator and can accurately detect instances of unconscious or inadvertent behaviors;
- The proposed method represents an expansion of the existing research on accidents involving heavy machinery operators, with a particular emphasis on the development and application of the novel method in this field.

5.1 Methodology

This study aims to fill the gap in research focused on preventing unintentional accidents involving heavy machinery operators and proposes an active real-time monitoring system that pays attention to the driver's state in order to help them when they fall into some subjectively uncontrollable abnormal states. For example, driver distraction can be quickly identified, and the driver can be alerted, thereby restoring the normal driving state. The method of recognition used in this study is mainly supported by the built-in program of the MR device worn by the driver. In the process of program development, the information (such as coordinates) of the driver's head and eyes are detected by calling the "distance and visual sensors" that are built into the device. At the same time, the driver's hands and the surrounding environment are analyzed using external sensors, and the current state of the driver is judged from multiple angles in order to confirm that the driver is capable of safely operating the vehicle.

Figure 2 introduced the research methodology of the proposed method. The analysis from research method, system design and architecture, prototype development and implementation to system evaluation is as follows:

Research method: An active real-time monitoring system based on the driver's state is proposed. Specifically, the method uses 11 actions as criteria for system evaluation and employs MRTK to analyze data from three angles: the wearer's eyes, head, and hands. During the experimental stage, testers conducted indoor and outdoor tests of each action and further analyzed the results by means of random sampling.

System design and architecture: The system design and architecture of this real-time monitoring system include MR equipment, distance and vision sensors, and external sensors. The system uses MRTK to analyze data from three angles of the wearer's eyes, head, and hands and monitor the driver's status in real-time. The system can alert drivers to improve their driving behavior and reduce the likelihood of accidents involving heavy machinery operators.

Prototype development and implementation: During the prototype development and implementation phase, MRTK is utilized to analyze data from the wearer's eyes, head, and hands and established 11 actions as the system evaluation criteria. Testers conducted indoor and outdoor tests of each movement and further analyzed the results by random sampling. The resulting system can monitor driver status in real-time and alert drivers to improve their driving behavior.

System evaluation: During the system evaluation phase, the comparison between the results of indoor and outdoor experimental scenarios is made to draw a final conclusion. The study focused on slow-moving heavy vehicles in construction zones, with the potential to extend to sparsely populated roads as far as the law allows. To ensure the recognition rate, this research classified abnormal behaviors by identifying specific actions.

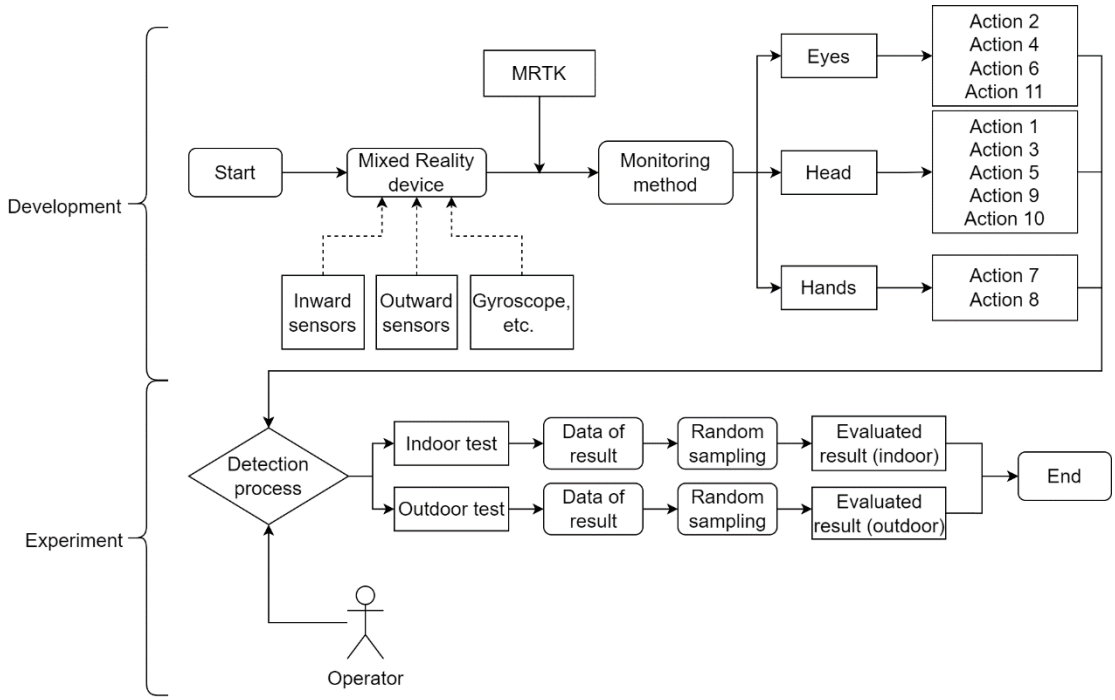


Figure 30 Research methodology.

5.1.1 Method introduction

Figure 31 demonstrates that fatigue, distraction, and inattentiveness can be detected and analyzed through various factors, including gestures, hand positions, head gaze position,

head movement, eye gaze position, and eye gaze status, among others, in heavy vehicle drivers.

Considering the potential risks posed by the dangerous and abnormal behaviors mentioned earlier, the monitoring activities will focus on three distinct areas, as outlined in the subsequent sections.

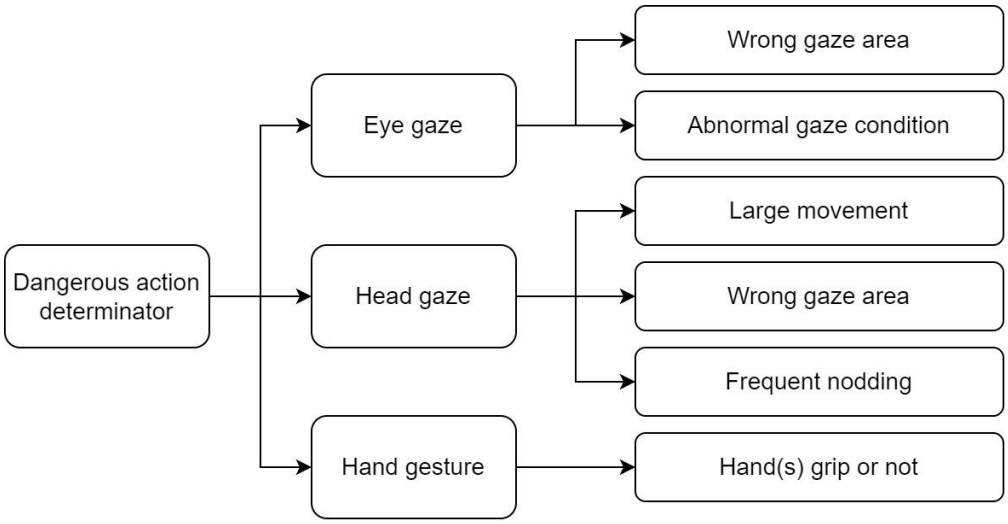


Figure 31 The main criteria for judging dangerous motions.

5.1.2 Hand-related recognition solutions

According to the introductory documentation of MRTK described in section 4.2.1, HoloLens 2 can recognize up to 25 nodes on one hand. By detecting the position of these nodes, MRTK can accurately recognize a wide range of hand movements within the visual range, making hand gestures one of the main inputs in MR interaction. Figure 32 shows that HoloLens 2's node recognition is relatively accurate, and several common input gestures that can be handled include grasping, tapping, air tapping, and long-distance cursor pointing.

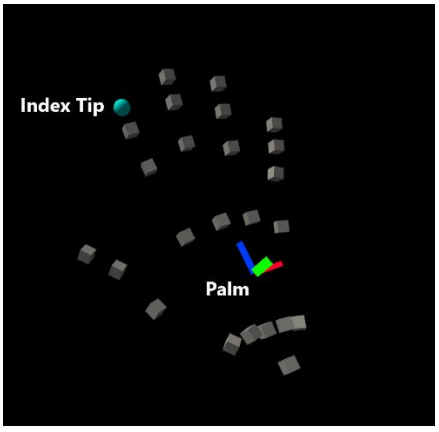


Figure 32 Hand nodes performance using HoloLens 2.

In this study, the operator's hand movements are an important benchmark for determining whether they are in an abnormal state (distracted or fatigued, etc.). These movements include releasing the steering wheel and exceeding a certain amount of time or taking the hand off the area where the steering wheel is located for too long. However, normal movements such as taking one hand off the steering wheel for a short period of time to operate the center control panel are not considered risky behavior. By comparing the similarity between actions, the research team found that the hand grip on the steering wheel is very similar to the basic grip in MRTK. Therefore, the default grip-me function of MRTK will be used in this study to simulate the operator's hand grip on the steering wheel. The current hand state of the operator can be quickly confirmed by the recognition of the grip action in the fixed area by the sensors of the MR device, as shown in Figure 33.

Figure 33 The “grabbing” motion recognized in MRTK is similar to holding a steering wheel.

Similar to the previous experiment, several hierarchical areas were set up within the driver's visual range in this experiment, including the “hand operation zone” around the steering wheel. During the real-time monitoring process, the presence or absence of the hand in the hand operation zone was used as an identification criterion to confirm the hazard level.

5.1.3 Head-related recognition solutions

The user's head can also interact with the system and receive feedback when the user's hands are engaged in a task and not idle. MRTK provides feedback on head pointing and dwell, which can help users to select or position points by head rotation alone.

In this study, the concentration is focused on operator state analysis to determine if they were driving dangerously or if there was a high probability that such behavior was about to occur. Typically, during prolonged distractions, the operator's head will unconsciously gaze at a fixed area, which is usually the policy operating area (e.g., looking at the curb, the rearview mirror, or the central control panel). When an

operator's attention is drawn to something external, depending on the state of the attractor, the operator may make a short, large head turn and not quickly return to the previous state. Also, fatigued driving can lead to involuntary operator actions, such as frequent head nodding, etc. Based on the above reasons, this research analyzed the operator's head direction and motion trajectory based on the MRTK head gaze and stay function and gyroscope data based on the previous experiment to ensure that this research can warn the operator in time when he/she is suspected or has been distracted and fatigued driving, so as to prevent accidents. By using MRTK, this research can contribute to improving road safety and reducing the risk of accidents caused by distracted or fatigued drivers.

5.1.4 Eye gaze-related recognition solutions

The HoloLens 2 is equipped with eye-tracking functionality, enabling users to select different objects within the same field of view without the need for head movements. This feature can be combined with gesture functionality to enable the selection, clicking, moving, and manipulation of virtual objects.

Our research indicates that a driver's eye state undergoes the most significant changes when they are either distracted or fatigued. Under normal driving conditions, a driver's gaze is primarily focused on the road ahead, with occasional glances to the rearview mirror or instrument panel. However, when a driver becomes distracted, their gaze may shift to unexpected directions or objects. Prolonged staring at distractions can have serious consequences, such as causing the driver to subconsciously steer in the direction of their gaze, potentially leading to collisions with oncoming traffic, pedestrians, or objects within a construction site. In addition, when a driver is fatigued, their eyelids may become heavy, their eyes unable to focus, and their blinking frequency may increase, necessitating more time to refocus after each blink. This state can impair the driver's ability to assess road conditions, leading to cognitive biases that could result in serious accidents or other consequences.

Based on the above, the driver's eyes can be monitored in real-time using the eye-tracking function in MRTK. By setting up sensing functions in areas where the driver should not be staring, the system can provide multi-dimensional early warnings and reminders to drivers who are staring at these areas while driving, immediately alerting them to return to a normal driving state and take sufficient measures such as decelerating. The eye-tracking system can also detect changes in the eyes' focus to identify erratic eye movements indicative of fatigue and promptly issue a warning. By analyzing the duration of time that the eyes are open versus closed, the system can determine whether the driver is too tired to continue driving. By using these methods, the system can determine whether the driver is fatigued, and the warning system can alert the driver to take action or stop for a rest to ensure their safety and prevent accidents.

5.2 Experiment

In the experiment, the latest version of Microsoft's MR headset is used, the HoloLens 2, as well as Logitech's G923 analog steering wheel controller (*Logitech,* n.d.). On the software side, Unity 3D and MRTK are used for the development of the MR environment, while Euro Truck Simulator 2 was used to construct the virtual driving scenarios.

Released in 2019, HoloLens 2 improves upon the previous version by expanding the viewing range two-fold, with a holographic resolution of up to 2048×1080 pixels per eye, making the projection of virtual objects clearer and presenting more interior details. HoloLens 2 has four visible-light cameras for head tracking, two infrared cameras for eye tracking, a 1-megapixel time-of-flight depth sensor for detecting depth information, an inertial measurement unit (IMU) sensor, and a digital camera capable of capturing 8-megapixel photos and 1080p 30-fps videos.

As the hardware foundation of this research, the HoloLens 2's advanced internal and external cameras and sensors enable functions such as hand recognition, head gyroscope, and eye tracking, which can realize the active monitoring of driving behaviors and perform multi-dimensional prediction of abnormal and dangerous behaviors in drivers. During the monitoring process, the MR device can issue visual warnings for some suspected dangerous driving behaviors at levels that will not affect normal driving, as well as audible auxiliary reminders. HoloLens 2 used in this experiment is shown in Figure 5.4.

In terms of human–computer interaction, HoloLens 2 can recognize all the joints of the hands and consequently be operated directly by the user's bare hands. It also supports real-time eye tracking.

The version of Unity 3D used in this simulation experiment was 2019.4.21f1, and that of MRTK was 2.6.0. MRTK for Unity is an open-source, cross-platform development kit for MR applications. It provides a cross-platform input system, basic components, and general building blocks for spatial interaction, including but not limited to camera modules, coordinate system, head gaze, text, hand and eye tracking, voice input, spatial mapping, and spatial sound effects.

5.2.1 Originalities and improvements

The aim of the present work is to consider the unrealized functions and the outstanding issues in the previous work, to further refine the existing questions surrounding the abnormal state of drivers and implement new functionality using the existing conditions.

First, the detection of drivers' hands has been improved. In the previous experiment, the detectable motions were limited to the grasping condition of the hands. Now, it is possible to detect which zone the hands are located in as well as determine the level of an emergency. It is also possible to avoid misjudging the driver's state and ensure the best response state of the driver during driving.

Second, the detection of drivers' head gaze has been improved. In the previous experiment, only the head-pointing area could be detected. Now it is possible to detect wider head rotation based on accelerometer data and vision, and thus the driver's abnormal behavior can be detected, and feedback can be provided more quickly. Meanwhile, a gyroscope-based method for monitoring nodding has been added to provide more accurate and timely warnings to drivers who are suspected of being fatigued.

Third, the detection of drivers' eye gaze has been improved. In the previous experiment, only eye gaze pointing could be detected. Now it is possible to monitor the state of the eyes. For example, when the movement of the eyes becomes erratic, the proposed method determines whether the abnormal behavior is dangerous. By recording the focus frequency of the eyes and monitoring whether the displacement of the focus point changes beyond a certain range within a fixed time, the state of the driver is determined, and an early warning can be immediately issued to prevent a potential accident. Another example is the problem of frequent eye closure caused by fatigue. By combining data such as whether the eyes can be detected and the ratio of open/closed eyes within a fixed time, the driver's state of alertness can be determined, and a warning can be immediately issued to remind them to drive safely.

5.2.2 Experiment setup

The model was developed in Unity 3D and later tested in HoloLens 2 to evaluate the actual MR experience. A virtual truck driving scene is created in Unity to simulate the field of view of a real truck driver, and this team built the overall framework of the early-warning system, which includes the location of surrounding environmental content and distance information in the virtual environment. As shown in Figure 34, this experiment simulates a visual model of a truck driver from the perspective of a right-hand steering wheel. Because the actual virtual environment is curved, some distortion may occur in the outer parts of the image. Based on driving habits and the layout of various parts of the vehicle, the driving field of vision is divided into the following four zones: gaze danger zone (red), gaze attention zone (yellow), gaze safe zone (green), and hand operation zone (blue). As mentioned in the previous sections, staring or hand movements in the wrong zone of the virtual environment will be judged as distracted or fatigued driving, and appropriate early warning measures will be activated.

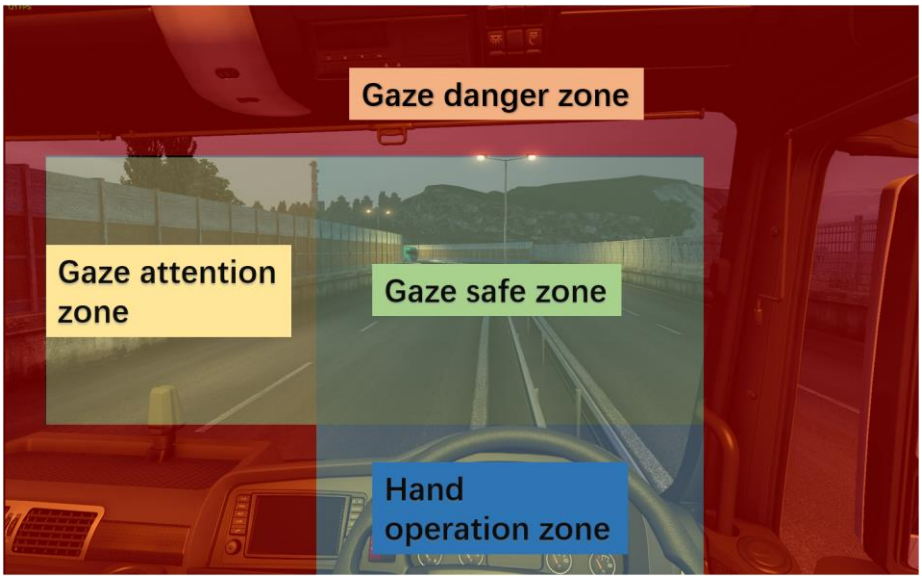


Figure 34 Different visual areas based on the operator's habit.

In this experiment, two-dimensional (visual and auditory) feedback is added to the early warning system in order to provide double redundancy for the driver's danger perception and to help them respond quickly in dangerous situations to avoid accidents. The visual and auditory warnings in this experiment were designed to be noticeable but not overly disruptive while driving.

Table 5 Experiment scenarios and action details.

Indoor/Outdoor	Eyes	Action 2	Keep the head still and eyes on the danger zone **(Red)**
		Action 4	Keep the head still and eyes on the attention zone **(Yellow)**
		Action 6	Keep the head still and eyes on the safety zone **(Blank)**
		Action 11	Mimic a fatigued state, with eyes closed or unable to focus.
		Action 1	Randomly point the head to the danger zone **(Red)**
	Head	Action 3	Randomly point the head to the attention zone **(Yellow)**
		Action 5	Randomly point the head to the safety zone **(Blank)**
		Action 9	Mimic a state of faigue, frequent nodding the head
		Action 10	Mimic a state of distraction, with a large swing of the head to the left or right
	Hand	Action 7	Look at the hands, fist, raise the head and lower the head in a few seconds.
		Action 8	Look at the hands, fist, raise the head, relase the hand and lower the head in a few seconds.

In this experiment, 11 actions have been defined to simulate the possible behavior of an operator in an unconscious state, such as when they are very tired or distracted. Among these 11 actions, 5 of them are related to eye movements, 4 to head movements,

and 2 to hand movements. The specific details of these actions are described in Table 5.

This experiment invited 4 testers, all of whom were young people between the ages of 20 and 30, due to the requirements of MR devices for certain qualities of the test subjects (understanding of MR devices, the sensitivity of eyes, etc.). In order to ensure the objectivity of the test, the 4 testers were divided into 2 men and 2 women. The experimental scenes are shown in Figure 35.

Figure 35 The experimental snapshots of testers.

The venue for this experiment was arranged in the author's laboratory and divided into two scenes: indoor and outdoor. During the experiment, environmental factors were kept consistent with minimal changes. The indoor scene was a windowless, temperature-controlled environment with a temperature range of 15-20 degrees Celsius and illuminated with reading lights. The outdoor scene was an open balcony environment with no significant obstructions, conducted in clear weather in the afternoon with temperatures ranging from 12-17 degrees Celsius. The experimental results were obtained from real machine testing using HoloLens 2 and did not use PC software to simulate data.

In the experiment, the main method of data collection is repeated observation and result recording of the testers and the experimental scene. To ensure the validity of the experimental data, as mentioned earlier, during the experimental process, this research try to keep other factors within controllable range except for the testers and designated test actions, such as environmental factors, instructions, and equipment.

Regarding the size of the data sample, in this experiment, each tester will perform 11 action instructions with 100 repetitions of each action, resulting in a total of 4,400 times. The data results will be presented in tabular form.

Since the uniformity of the testing efficiency of testers cannot be guaranteed, and outdoor experiments are included, various outdoor factors cannot be reasonably controlled. Therefore, the direct results of this experiment will have some errors that may affect the final experimental conclusion. Hence, multiple random sampling methods will be used to randomly select 80 sets from the 100 sets of results for each person and each action and re-analyze them. This way, individual bad results in the 100 sets of data can be avoided from affecting the accuracy of the overall result as much as possible.

5.2.3 Development concept

Some components from the MRTK are used to implement some methods, including basic hand recognition and motion feedback. For example, because of the high similarity between the action of grasping the steering wheel and the action of virtual grasping, the *NearInteractionGrabbable* and *ManipulationHandler* components are used to identify whether the driver is holding the steering wheel correctly. In this study, a script named *HandDetector* (Figure 36) is also written originally, which is used to recognize the 3D coordinates of the hands in order to detect when the hands leave the operating area and issue an early warning alert.

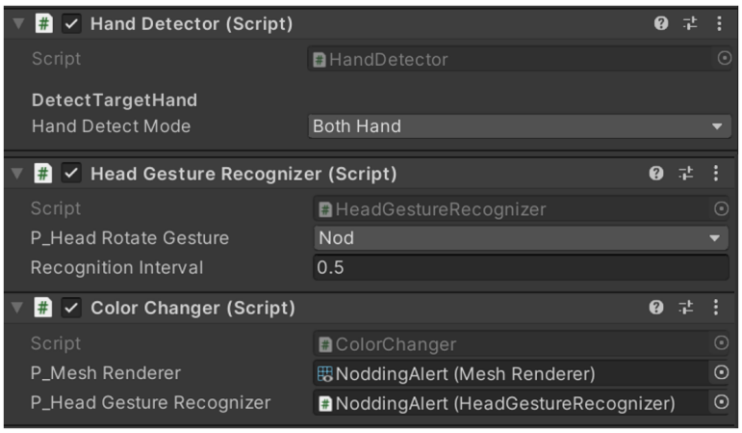


Figure 36 New-added scripts in the experiment.

Some functions in the MRTK are utilized to detect the angle of the driver's head and the focus of their eyes. For example, *EyeTrackingTarget* is used to lock the location of the eye focus, while functions such as *GazeProvider* and Gaze are used to activate the internal sensors of the MR device, thereby realizing eye tracking and information collection. When the driver is staring at an area within visible range, different colors will be displayed according to the level of danger (the safe zone is transparent). When the driver's eyes or head point toward a danger zone or attention zone and remain there

for more than 0.8 s, this zone will turn translucent red or yellow as a visual warning to alert the driver. At the same time, in the gaze danger zone (displayed in red), because of the high level of danger, first a visual alarm is triggered, followed immediately by an audible alarm, while in the gaze attention zone (displayed in yellow), the auditory alarm will be activated later in order to reduce the level of interference. If the driver quickly resumes normal driving after receiving the visual warning, the audible alert will not be triggered to avoid disturbing the driver.

In this experiment, two new scripts are implemented to realize the detection of frequent nodding during fatigued driving: *HeadGestureRecognizer*, and *ColorChanger* (Figure 8). By collecting gyroscope data, this team was able to identify the motion of the head in three dimensions about the x-axis, as shown in Figure 36, and thereby realize the detection of frequent nods. During the experiment, this team found that when the driver nodded frequently due to fatigued driving, the head also dropped significantly and the target point of the front of the head would fall into the hand operation zone. Therefore, to further improve the accuracy of fatigued driving detection, a function is added to the hand operation zone that identifies the dangerous area level. When the head falls within this zone, the two-dimensional early warning system will be triggered to alert the driver and help them correct their status to avoid an accident.

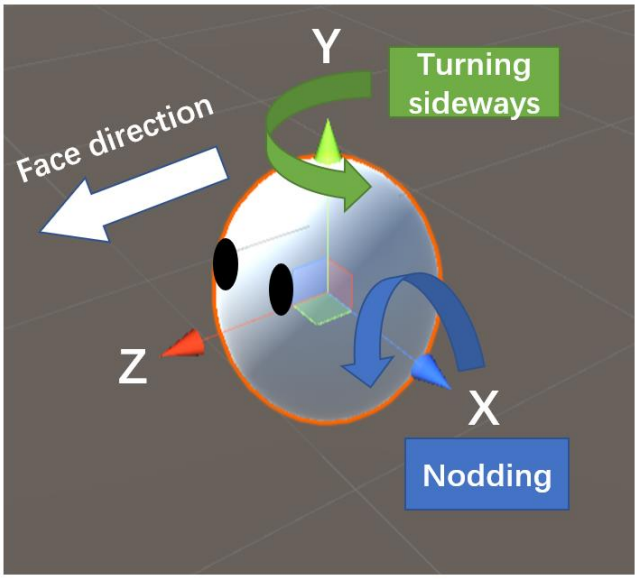


Figure 37 3-axis gyroscope and head movement.

While driving, it is dangerous for a driver to become distracted by sounds or other external stimuli and turn their head sideways to observe them. To realize the recognition of such a large head rotation, the three-axis gyroscope inside the HoloLens 2 was utilized to detect large changes in the y-axis, as shown in Figure 37. This process allows the driver to be alerted in a timely manner, helping them return to normal driving conditions.

5.3 Result

Experiments are tested on a HoloLens 2 in a real environment. Since the use of MR equipment to take photos and records will deepen the transparency of the virtual module when the results are presented, the real machine display screen will be more transparent than the pictures shown in this article (it will not affect the basic operation).

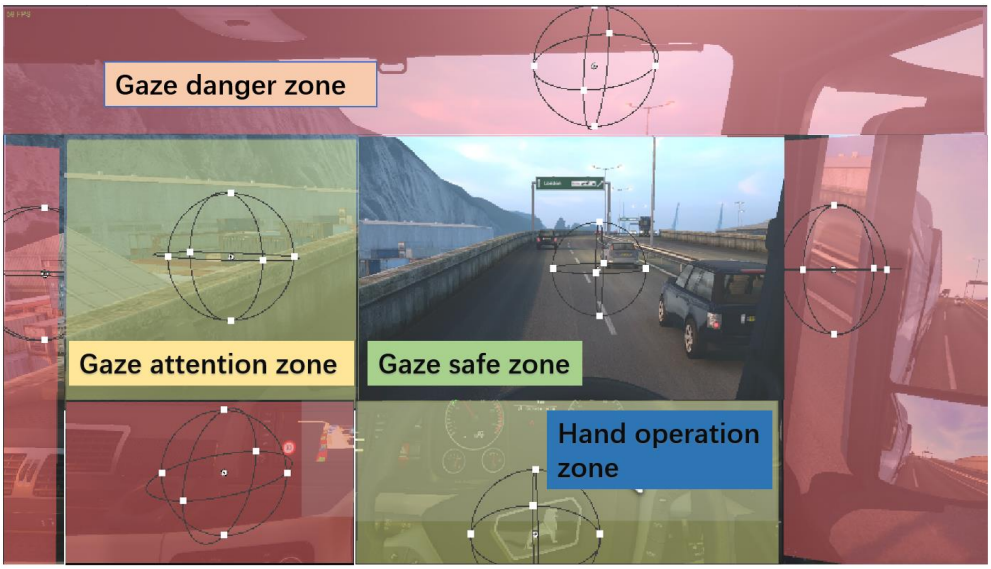


Figure 38 Warning results during distracted driving. (head/eyes)

The results for head pointing and eye pointing are generally similar, and thus, they are grouped together in the results. In the cockpit environment of the simulated truck shown in Figure 36, when the tester's head (eyes) points to each zone, different visual and auditory feedback was obtained according to the danger level of the zone. The angle of view in this figure was obtained when the head was not moving. Because the eye gaze is invisible, it was marked with a three-dimensional ball for better observation. By including the hand operation zone in the warning area, this team was able to improve the accuracy of detecting fatigued driving and effectively reduce the likelihood of an accident.

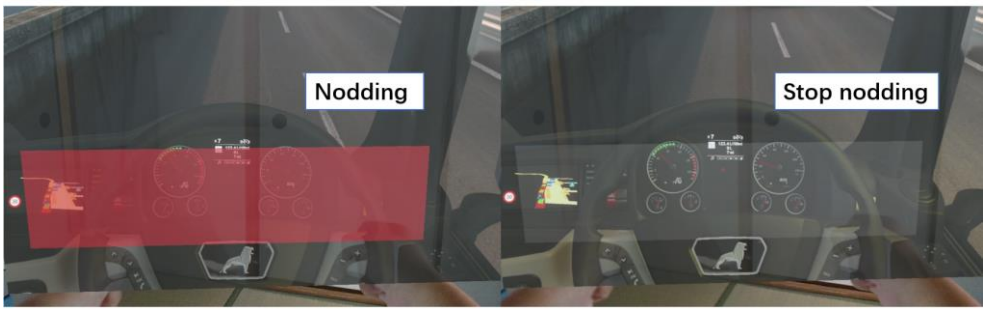


Figure 39 Frequent nodding detected during fatigued driving.

Figure 39 shows the detection of frequent nodding movements during fatigued driving. The driver's head will fall unconsciously when they are in a fatigued state. A warning sign can be placed near the hand operation zone without affecting the normal driving angle of view and the warning sign will be triggered under conditions such as when the driver makes a big nod or nods constantly. When the driver returns to a normal driving state, the warning sign will instantly become transparent again. In addition, large head movements such as swinging caused by distracted driving can also be addressed according to the same principle. When a large head swing is detected, warning signs on the side of their field of view will alert the driver to return to a normal driving state.

Figure 40 Upgraded detection of grabbing motion.

Table 6 Real machine test data recording (original data).

Original actions (100 times)		Male 1 (M1)		Male 2 (M2)		Female 1 (F1)		Female 2 (F2)		M1	M2	F1	F2	Average
		√	×	√	×	√	×	√	×	Accuracy				
Indoor	Action 1	89	11	94	6	92	8	92	8	89%	94%	92%	92%	92%
	Action 2	92	8	93	7	94	6	94	6	92%	93%	94%	94%	93%
	Action 3	96	4	95	5	91	9	92	8	96%	95%	91%	92%	94%
	Action 4	97	3	97	3	98	2	96	4	97%	97%	98%	96%	97%
	Action 5	99	1	98	2	99	1	98	2	99%	98%	99%	98%	99%
	Action 6	99	1	97	3	99	1	99	1	99%	97%	99%	99%	99%
	Action 7	91	9	89	11	92	8	88	12	91%	89%	92%	88%	90%
	Action 8	96	4	92	8	97	3	91	9	96%	92%	97%	91%	94%
	Action 9	90	10	88	12	91	9	87	13	90%	88%	91%	87%	89%
	Action 10	85	15	83	17	82	18	86	14	85%	83%	82%	86%	84%
	Action 11	88	12	84	16	84	16	87	13	88%	84%	84%	87%	86%
Outdoor	Action 1	85	15	91	9	89	11	90	10	85%	91%	89%	90%	89%
	Action 2	89	11	90	10	91	9	91	9	89%	90%	91%	91%	90%
	Action 3	95	5	91	9	90	10	91	9	95%	91%	90%	91%	92%
	Action 4	95	5	93	7	95	5	93	7	95%	93%	95%	93%	94%
	Action 5	96	4	96	4	97	3	97	3	96%	96%	97%	97%	97%
	Action 6	95	5	96	4	98	2	95	5	95%	96%	98%	95%	96%
	Action 7	87	13	83	17	88	12	86	14	87%	83%	88%	86%	86%
	Action 8	93	7	89	11	91	9	89	11	93%	89%	91%	89%	91%
	Action 9	89	11	85	15	90	10	84	16	89%	85%	90%	84%	87%
	Action 10	87	13	82	18	83	17	84	16	87%	82%	83%	84%	84%
	Action 11	82	18	80	20	83	17	87	13	82%	80%	83%	87%	83%

Figure 40 shows the results of detecting the grabbing motion during driving. In this figure, an update of the locator indicator of both hands is also presented, which will disappear after the driver maintains the grab state, thereby reducing the impact on the driver's normal driving state.

Table 6 presents the recognition results and accuracy information obtained from real-world testing using HoloLens 2 in both indoor and outdoor environments. The tester was situated within the test environment and wore MR equipment in a normal fashion. Upon system initiation, the tester followed the observer's instructions step-by-step and recorded the outcomes. The experiment was evaluated using a standard in which the experimenter would be marked with a (√) if they received correct feedback within the specified time frame (normally less than 3 seconds) after completing a specified action, and a (×) if they did not. For each action, the average accuracy was calculated and evaluated. In this experiment, each person tested each action 100 times, a total of 4,400 times. In the table, correct behavior recognition and the corresponding average recognition rates are indicated. Correct recognition with an accuracy of more than 90% is marked in green, while those with an accuracy of less than 90% are marked in yellow. Similarly, average recognition rates of more than 90% are marked in green, and those of less than 90% are marked in yellow.

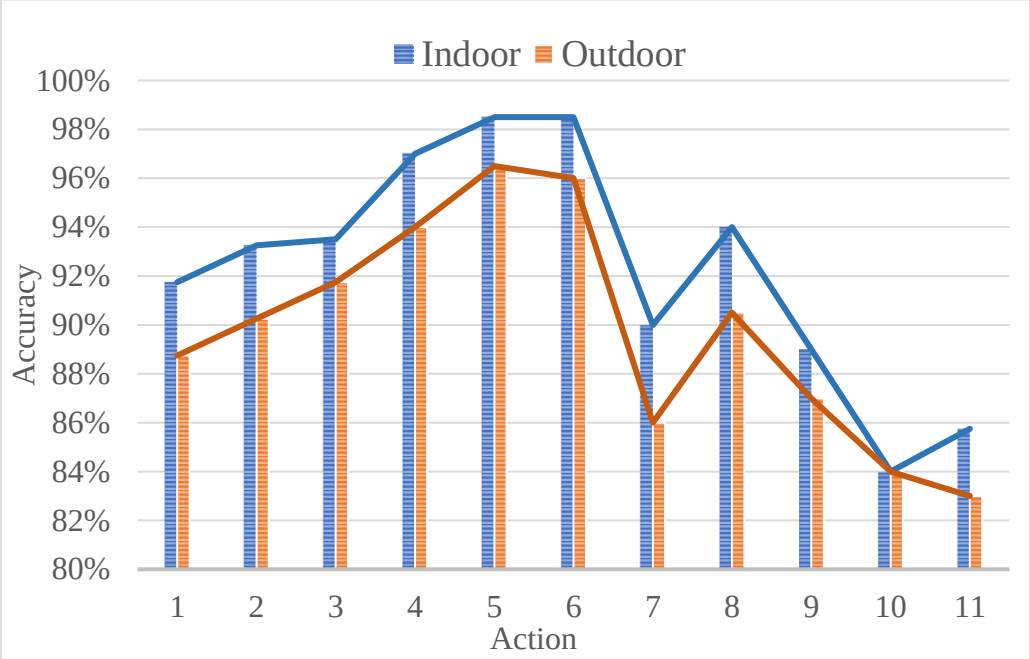
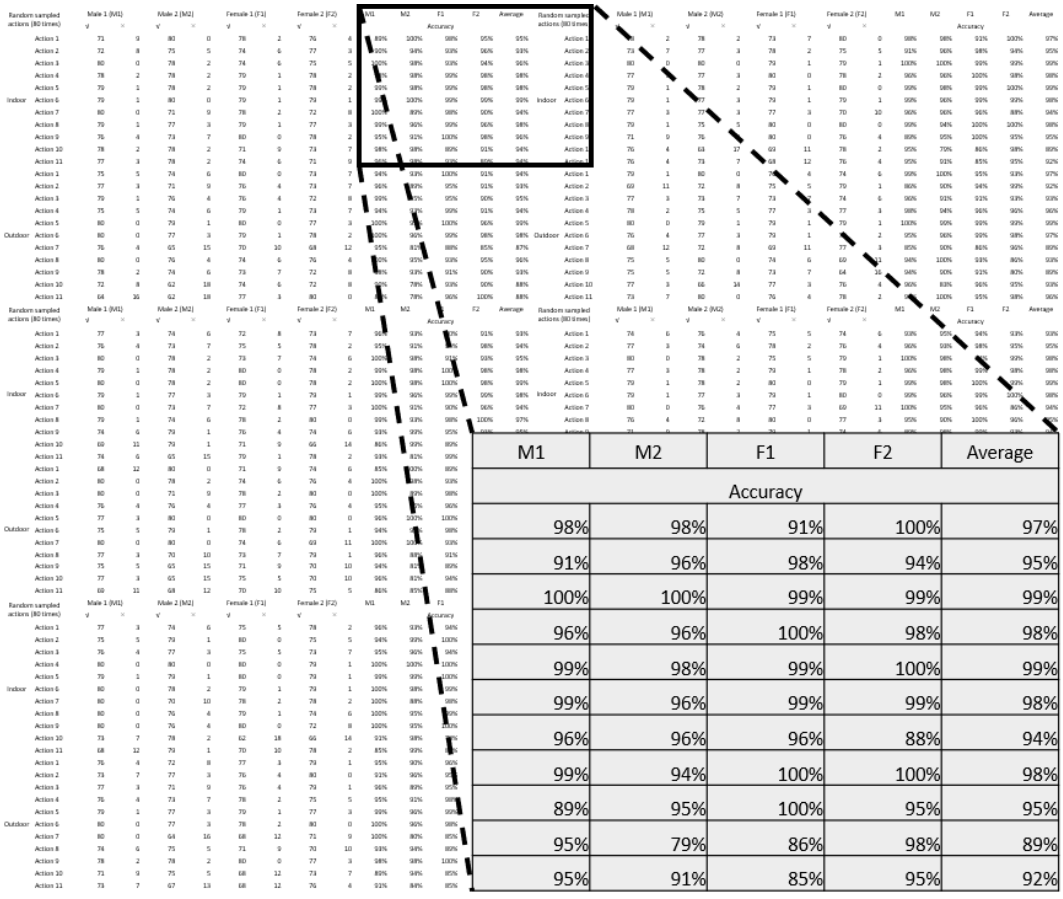


Figure 41 Trend chart of in/outdoor accuracy under the original data.

Figure 41 depicts the accuracy trend of indoor and outdoor experiments using the original data. The figure clearly illustrates that the monitoring and recognition accuracy of actions remains above 80% for both indoor and outdoor settings. Moreover, the recognition accuracy of individual actions is almost close to 100%. However, due to

the impact of outdoor environmental factors, the recognition accuracy of almost all actions outdoors is consistently lower than that of indoors.



M1	M2	F1	F2	Average
Accuracy				
98%	98%	91%	100%	97%
91%	96%	98%	94%	95%
100%	100%	99%	99%	99%
96%	96%	100%	98%	98%
99%	98%	99%	100%	99%
99%	96%	99%	99%	98%
96%	96%	96%	88%	94%
99%	94%	100%	100%	98%
89%	95%	100%	95%	95%
95%	79%	86%	98%	89%
95%	91%	85%	95%	92%

Figure 42 Partial data in multiple random sampling.

However, analyzing data solely based on raw data may result in some inaccuracies. For example, during the experiment, there may be some extreme data points that could cause an increase or decrease in the overall recognition rate. To further enhance the objectivity of the results, the method of multiple random sampling is applied to process the raw data. As shown in the example in Figure 42, 80 sets of data are randomly selected from the 100 sets of data for each action multiple times, thereby reducing the impact of extreme data on the overall trend.

Figure 43 shows the trend graph obtained for the accuracy data after multiple random sampling. By this step, most of the extreme data in the original data that affects the overall trend is avoided. From the figure, it is known that after random sampling, the accuracy of both indoor and outdoor data reaches more than 85%, which is an overall improvement over the accuracy of the original data. At the same time, after random sampling, the overall accuracy derived from the outdoor data remains lower than the indoor accuracy. The specific reasons and the analysis will be explained in the next section.

Figure 43 Trend charts of in/outdoor accuracy after multiple random sampling.

5.4 Summary

5.4.1 Discussion

In this study, a MR-based active hazard prevention system for heavy machinery operators is presented. The approach details the development and experimentation of an active real-time monitoring system that is specifically designed to prevent accidents involving heavy machinery operators. As shown in Figure 14 and Figure 16, the proposed method does not depend on the subjective behavior of the operator and can accurately detect unconscious behaviors that may lead to hazards during operation. Furthermore, the proposed method represents an expansion of existing research on accidents involving heavy machinery operators and represents a new research direction and a unique approach in this field.

The hardware is used to monitor several major critical parts of the operator (hands, head, and eyes) in real-time for safe driving. As shown in Figures 11-13, the sight distance is divided into different danger level areas while ensuring that safe driving is not compromised, and the driver status is analyzed with real-time feedback. At the same time, MR devices are effectively used to prevent accidents by analyzing the main causes of accidents among site drivers and linking these causes to monitorable driving behaviors and body parts.

Despite hardware limitations, sensors are fully utilized to present experimental results without the need for additional connections or calculations. The local device provides all the computational power needed.

In previous MR-based construction-related studies, the focus has usually been on what the users of the method can obtain from the external environment (e.g., pre-construction learning and training, etc.). This study, however, starts from the opposite direction and focuses on what information can be obtained from the users themselves and how this information will affect the external environment. Meanwhile, in previous studies on personnel safety in the construction industry, the physical state of personnel is mostly analyzed from a biological perspective, requiring the use of complex external equipment, such as electrical signal sensing, which is difficult to apply effectively in the real world. Other research has included adding various sensors to heavy machinery or vehicles to improve safety, but many of these features rely on subjective judgment by the operators themselves. In contrast, in this study, operator behavior and habits are monitored and analyzed in real-time and can be identified and detected regardless of whether the operator is in a subjective and controllable state, thus further ensuring safety and avoiding potential accidents.

The results of this study show that the proposed method has a high recognition accuracy rate and can maintain a high recognition success rate even when detecting various actions for a long time. It is worth noting that in the abnormal detection of the head and

eyes (actions 1 to 4), the final recognition rate is above 90%. According to the later recall of the testers, most of the recognition errors are due to gaze to the corner of the field of view of the MR device, and the edge of the sensing area. In the anomaly detection for hand grasp (actions 7 and 8), there is a large gap between indoor and outdoor results, and the accuracy of grasp detection is lower than that of let-go detection, mainly because in grasp detection, Since the field of view of the MR equipment is too small when the operator raises his head, his hand is out of the field of view of the sensor, so when the operator looks back, the sensor will make more mistakes in the judgment. In the abnormal detection of large head rotation and rapid nodding (actions 9 to 11), the detection of eye focus in the outdoor environment is significantly lower than that in the indoor environment, mainly because the outdoor environment is complex, and the introverted sensor is too much interfered lead to. As for the misrecognition caused by looking at the normal operating angle (actions 5 and 6), most of them are misjudgments caused by habitual head turning and eye shifting. In total, under the given detection conditions, the proposed method can identify specific actions caused by fatigue or distracted driving that may occur in various driving processes with high accuracy and without delay, enabling real-time monitoring during normal operation and protection.

For future expansion, the application of the proposed approach to the target industry has also been considered: construction. Although real-world testing has been conducted, it has been done in a relatively ideal environment and has not actually been introduced into heavy machinery or vehicles. Therefore, in order to be able to adapt to the complex environment of the construction industry, targeted development for different machines or vehicles is essential. At the same time, the proposed method has a lot of room for improvement. In this experiment, it is known that the need to identify the eyes makes it troublesome to calibrate it every time the tester is changed. At the same time, due to hardware limitations, errors are often made when detecting the edge area of the field of view, and the detection of the four corners is sometimes less accurate. Therefore, in future work, overly intuitive area edges should be avoided, while reducing the information in corners and other locations and placing more content in more easily perceived areas. Regarding technical limitations, first, MR devices are still not mainstream products and therefore costly to use widely. Second, although HoloLens 2 doubles the visual range compared to its predecessor, it is still too small for holographic displays, so if you want to adapt to the field of view of heavy machinery, for example, you need to turn your head more significantly, which is quite inconvenient. Then, because HoloLens 2 is too integrated, so poor heat dissipation, and long-term operation after the hologram will be torn, the system slows down and other problems, while the long and efficient operation of HoloLens 2 battery life will also be shortened. Therefore, if the method needs to be applied to the construction industry in the future, it may be necessary to meet the basic functions of flat replacement products.

In order to effectively utilize this method in construction sites, it is imperative to optimize the equipment in accordance with the unique features of the work environment

and scenarios. One such optimization measure is to integrate sensors, such as temperature and humidity sensors, within the heavy machinery. This will enable an expanded dimension of monitoring the operator's condition, which will ensure a higher level of detail in assessing the operator's state of wakefulness and comfort.

Meanwhile, to ensure the successful implementation of this method, it is essential for the implementing organization or unit to enhance operator training and technical support. This will enable the operators to properly wear and comprehend the MR device, as well as to promptly return to the normal driving status in accordance with the guidance provided by the system. Moreover, in cases where heavy machinery operation is complex or hazardous, this method can be employed as a necessary safety measure to safeguard the lives of the workers. Furthermore, in expanding this method to sparsely populated roads, it is crucial to optimize and adapt to the various road conditions to ensure accurate identification of the driver's condition and environmental changes.

5.4.2 Conclusion

In this study, an MR-based advanced active monitoring system for heavy vehicle drivers is proposed. The purpose of it is to monitor operators in real-time to avoid dangerous behaviors despite of whether he or she is conscious or not. The MR device (HoloLens 2) is used as the hardware, simultaneously calling and collecting information from various sensors to monitor the operator's state in real-time, so as to ensure that he will not fall into an unconscious dangerous state. This approach is novel in that it focuses on the active provision of early warnings based on the operator’s state. By using HoloLens 2 to monitor several body parts of the driver in real time, the system can confirm whether the driver is in a state of distracted or fatigued driving. This study is based on the hands, head, and eyes—the three main body parts used when operating; by analyzing the state of the operator based on abnormal behaviors detected in these body parts, early warning feedback can be provided to the operator, thereby helping to avoid an accident. In this study, the following points are mainly achieved:

- The development of an active real-time monitoring system is introduced specifically designed to avoid accidents or incidents caused by the dangerous behavior of heavy machinery operators.
- The proposed method can achieve real-time and accurate detection of unconscious or unintentional behavior independent of the subjective behavior of the operator.
- The presented method represents an extension of existing research on accidents involving heavy machinery operators, with particular emphasis on the development and application of new methods in this field.

From the analysis of the results, it is known that some trend graphs were very typical, therefore further analysis is applied. In most cases, it can be concluded from the data that indoor recognition accuracy is generally higher than outdoor recognition accuracy.

However, in some cases, the outdoor recognition accuracy of some actions is higher than the indoor recognition accuracy. Therefore, this research reasonably speculate that although changes in environmental factors may lead to a decline in recognition accuracy, this is not an inevitable factor. Therefore, the future improvement of this method, which is mainly aimed at outdoor work, may result in better results than the experimental environment, which also provides a theoretical basis for the future implementation of this method.

In future improvements, the concentration will be solving the hardware deficiencies of MR equipment and reducing misjudgments and errors caused by viewing angles. At the same time, the current mainstream MR equipment will be selected and compared to ensure that when it is applied to the construction industry in the future, it can reduce conflicts at the software and hardware levels, thereby improving efficiency.

Chapter 6

Discussion

6.1 Brief discussion

The study titled "Research on construction safety techniques via multi-sensor information fusion" is a comprehensive exploration of how information technology can be leveraged to improve human safety on construction sites. The practical significance of this study cannot be overstated, as worker safety is a crucial aspect of any construction project.

The study takes a targeted approach by focusing on the two main occupations on construction sites: workers and large machinery operators. By considering the unique characteristics of each occupation, the research plan is tailored to maximize accuracy and effectiveness. For instance, to deal with the uncertainty of workers' positions and movements, a dual-sensor method of remote and wearable devices is used, while head-mounted devices are used for machinery operators who have fixed positions but need to frequently operate hands and feet.

The equipment selection and experimental design are also carefully considered, with cost-effectiveness being a primary concern. The study uses relatively affordable depth cameras and IMU sensors that are suitable for small to medium outdoor scenes, which can significantly reduce deployment costs across different construction sites. The hardware's basic functions also ensure stability and upward compatibility, making it easier to upgrade the products.

Traditionally, construction safety discussions have focused on the building or environment itself, but this study shifts the focus to the workers who are working on the site. Despite significant advances in the construction industry over the past decade, worker safety has remained a significant issue. By focusing on human safety, this study

aims to draw attention to the industry's safety shortfalls and provide a solution to ensure the most basic security issues are guaranteed.

Therefore, "Improving the safety of construction site personnel using multi-sensor data fusion" is a well-planned, practical study that addresses one of the most critical issues in the construction industry. The study's results could revolutionize the industry by providing practical, cost-effective solutions to ensure worker safety.

While the majority of the experiments in the "Improving the safety of construction site personnel using multi-sensor data fusion" were conducted in a controlled laboratory environment, steps were taken to add influencing factors to make the experiments more realistic. As such, the study's conclusions are oriented and informative, providing valuable insights into how information technology can be used to improve human safety on construction sites.

One of the key findings of the study is that the dual-sensor method of remote and wearable devices is highly effective in tracking worker movements and ensuring stability without interfering with their work. This approach can significantly improve worker safety by alerting them to potential hazards and reducing the risk of accidents. The study also found that head-mounted devices are effective in tracking large machinery operators' movements and reducing the direct impact on them. This technology can help reduce the risk of injuries caused by machinery accidents, improving worker safety and reducing the risk of equipment damage.

The study concludes that information technology, when used in combination with sensors and data fusion techniques, can significantly improve worker safety on construction sites. While the experiments were conducted in a laboratory environment, the study's conclusions provide a solid foundation for future research and the development of practical solutions to improve construction site safety.

Overall, while the study's experimental content was based on a research laboratory, the conclusions were enriched by the addition of influencing factors that made the experiments more realistic. The findings highlight the potential for information technology to improve worker safety on construction sites, which could have a significant impact on the industry's safety practices.

6.2 Reflecting of results

For each experiment, the research added a discussion or introduction section at the end of the chapter, and in this section, the results of these three experiments will be summarized and discussed.

6.2.1 Experiment 1

In the experiment on worker safety, this research discussed the recognition of actions performed by workers on construction sites using a fusion of signals from depth cameras placed at a distance and IMU sensors worn on the body. This study is an extension of a previous study, which not only discusses the detection of up to 10 actions on construction sites but also extends the range of depth cameras recognizable by combining different signals. The method achieves similar detection accuracy even beyond the optimal detection distance, making it suitable for environments in more medium-sized construction sites.

In the first experiment, this research proposed a depth-assisted method called "Selective Depth Inspection (SDI)" which divides the recognition process into two steps to solve the difficulty of depth recognition in complex environments. The first step is to detect whether there is a person in the monitored environment. If a person is present, the second step is selected to recognize the area where the person is located and perform depth recognition and optimization. This method is more effective in detecting objects outside the optimal detection distance of the depth sensor while reducing the computational pressure on the depth camera during long-term detection.

From the comparison of the results, the use of the SDI method enables long-distance motion detection to reach the accuracy of short-distance detection at the same recognition distance (beyond the optimal detection distance of the depth sensor), and the lack of depth information is effectively compensated by image processing.

Our experiment shows that the average accuracy of motion recognition by multi-sensor fusion at short and relatively long distances, assisted by optimization from the SDI method, can reach 93.27% and 92.80%, respectively. Compared with the results without using the SDI method at the same distance, the accuracy has improved by about 12%.

In the experiment, the action model data used was mostly obtained by adding white noise or other influencing factors to the standard data. As a result, in future real-world experiments, the performance may be relatively lower due to the increased uncertainty resulting from additional unknown factors.

6.2.2 Experiment 2

In this experiment, this research presented an innovative approach for actively monitoring the safety of drivers of large vehicles on construction sites, utilizing a mixed-reality device. This analysis of relevant research revealed that current safety methods for these drivers are mostly passive, relying on safety education and accident simulations, with feedback dependent on the driver's conscious reaction. However, accidents can often occur due to factors outside of the driver's control or unconscious operation, such as fatigue or distraction.

To address this issue, this research focuses on monitoring the hands, head, and eyes - three body parts commonly used during driving maneuvers. Using the MR device's sensors, this research monitor and analyze the driver's actions in real time, providing immediate feedback on their actions. The device is worn on the head to minimize any impact on the driver's normal operation and vision.

Our experiment successfully demonstrated the feasibility of the method, with no unrecognized states observed. Although this approach offers a solution to safety monitoring for drivers of large vehicles on construction sites, challenges remain regarding the poor portability of MR devices and the inconsistency of regulations across different regions. Addressing these practical challenges will be essential for implementing this approach in real-world settings.

6.2.3 Experiment 3

In this experiment, this research is expanding the research question to include operators of large machinery, building upon the previous experiment's focus on drivers of large vehicles on construction sites. This research has also broadened the scope of the investigation based on the results of the previous experiment. While previously only relatively simple states are monitored, such as hand grasping and head gaze. Now, this research aims to capture more complex situations that arise when an operator is fatigued or distracted. These include scenarios such as the operator removing their hands from the steering wheel or operation panel, dozing off due to excessive fatigue, and being distracted by certain objects leading to large head turns.

As this research has conducted more experiments, it has observed an increase in the complexity of the test phenomena. However, some limitations are also encountered. Specifically, due to the field of view of the MR device, it can be noticed that hand detection can be limited when the operator's head is in a higher position. As a result, this research will need to modify the methods to account for this limitation.

Chapter 7

Conclusions

7.1 Summary

The topic of hazard prevention on construction sites is of utmost importance, as it ensures the safety of workers and operators while working on site. One of the key aspects of hazard prevention is worker monitoring, which involves mid-range action recognition using visual and sensor information. In order to ensure worker safety, it is necessary to detect the movements of workers who are at a medium-to-long distance from the surveillance camera. This can be achieved through the extension of the recognizable distance of the depth camera. Additionally, it is important to identify the 10 kinds of actions that often appear on construction sites.

The second part of hazard prevention on construction sites involves operator monitoring, which focuses on the status recognition of the operator of the MR head-mounted device. The camera and sensors inside the device monitor the operator's status in real-time to identify the operator's intention. The main method is to use the internal camera to monitor the gaze position of the pupil, and at the same time use the orientation of the external camera to simulate the wearer's head orientation. By identifying the gaze position of the pupil, it is possible to determine the wearer's personal state at that time, such as whether they are distracted or fatigued. The external camera and depth camera can also recognize the wearer's hands and detect the actions at that time, such as pointing or grasping.

The third part of hazard prevention on construction sites builds upon the content of the second article by adding more recognition and identification standards for the simple recognition of eyes, heads, and hands. For example, the recognition of eyes now includes the ability to detect "slack eyesight, such as being unable to focus under certain circumstances." Additionally, the recognition of the head now includes the nodding action, which is necessary for identifying head pointing. Furthermore, after the wearer is attracted by something, the head may follow the attraction to make a large turn, which could cause possible danger. The recognition of the pointing of the head will be far

away from the established recognition area, thus highlighting the importance of this aspect of hazard prevention. The recognition content for hands has also been increased, such as whether the hands are in the specified area. It is important to note that the main reasonable areas for the presence of hands are the range near the steering wheel and a part of the center console area. If the hand stays in this area for a short time, it is reasonable, but if it stays in this area for a long time, it means that the driver's driving has been harassed, and the driving state at this time is also unsafe.

Overall, hazard prevention on construction sites requires a comprehensive approach that includes worker and operator monitoring. By using advanced technologies such as depth cameras and MR head-mounted devices, it is possible to detect potential hazards and prevent accidents from occurring. The continuous improvement and modification of recognition and identification standards will ensure that workers and operators remain safe while working on-site.

7.2 Conclusion

The research proposes three methods to address personnel safety issues on construction sites, with a focus on workers and operators. The first method aims to achieve timely early warning and rescue of blind spots in different construction situations through monitoring personnel, to ensure the safety of workers in complex conditions on the construction site. To achieve this, the researchers propose a Motion Recognition Method that uses a technique called Selective Depth Inspection (SDI) to optimize the computational pressure of depth maps and improve their accuracy. They also extend the detection distance of a normal depth camera and improve the depth camera-based motion recognition with SDI, which effectively reduces the detection errors and calculation burden of a broad range of depth data, while enhancing the recognition distance and accuracy of the depth camera in a selected area. Additionally, they use a portable computing terminal instead of a single depth camera to achieve local analysis, which avoids the computing burden caused by transferring a large amount of data to a central processing unit. The study also demonstrates that using different types of sensors to recognize human motion improves the accuracy of motion recognition.

The second and third approaches aim to achieve active monitoring of operators in a way that does not interfere with their normal work. This approach is unique because it focuses on providing real-time early warnings based on the operator's status. The methods rely on the MR device HoloLens 2 to monitor the operator's head, hands, eyes, etc. in real-time and provide visual and auditory feedback when an abnormal situation is detected, urging the operator to quickly recover to normal status. Although the overall accuracy of the system is high, its detection effectiveness in some cases needs further improvement due to hardware limitations such as field of view and battery life. The

study shows that MR-based active detection of heavy machinery or vehicle operators is highly feasible and yields effective feedback.

These methods propose innovative techniques to enhance personnel safety on construction sites and improve early warning and rescue systems. The study provides insights and solutions to address personnel safety issues and demonstrates the potential of emerging technologies in this field.

7.3 Limitations and future works

Limitations in Method 1: One limitation of this research is that it limited the identifiable actions on the construction site to only 10, which may not cover all possible actions that can cause injury to workers. Moreover, due to hardware limitations, the current SDI method cannot provide real-time detection, and the collected data and simulation results are analyzed over time. In practical applications, data collection and analysis are usually carried out in a short period of time, so real-time identification of specific actions is essential.

Furthermore, this method has only been tested in a simulated environment, and its accuracy may decrease when applied to a real construction site due to unforeseen disturbances. During the experiments, it is also noticed that some actions with similar points of interest had similar trajectories, leading to misjudgment. To improve accuracy, this research plan to consider the movement of interest points or explore other methods to more precisely define similar actions in future research. This research also aims to add real-time alarms based on motion recognition to the detection system, which can enhance the hazard prevention scheme on construction sites.

Limitations in Methods 2 and 3: The research focuses on the driver of vehicles and heavy machinery, and reducing their actual experience under available conditions is crucial. However, the HoloLens 2 device used in the experiments is not suitable for field application due to its weight and dark design to ensure holographic image clarity. This may cause discomfort to the driver and compromise their ability to concentrate for prolonged periods. Moreover, the high cost of HoloLens 2 presents a significant cost consideration.

Additionally, the method has limited adaptability to different types of heavy vehicles or machinery. To improve adaptability, the research plan to explore UI adaptation or self-programming techniques that align virtual space points and real vehicle parts to ensure adaptability to different types of vehicles or machinery. However, in some European and Asian countries, wearing MR equipment while driving on public roads is not yet legal, and this research must be careful not to violate local regulations while promoting this method.

In the future, the reliability and safety of emerging VR/MR devices should be considered, and this research must also address legal concerns before promoting this method. Overall, these limitations provide opportunities for further research and improvement to enhance the accuracy, efficiency, and feasibility of the methods.

References

Alsamadani, R., Hallowell, M., & Javernick-Will, A. N. (2013). Measuring and modelling safety communication in small work crews in the US using social network analysis. *Construction Management and Economics*, *31*(6), 568–579. https://doi.org/10.1080/01446193.2012.685486

Amine Elforaici, M. E., Chaaraoui, I., Bouachir, W., Ouakrim, Y., & Mezghani, N. (2018). Posture Recognition Using an RGB-D Camera: Exploring 3D Body Modeling and Deep Learning Approaches. *2018 IEEE Life Sciences Conference (LSC)*, 69–72. https://doi.org/10.1109/LSC.2018.8572079

Angell, L., Auflick, J., Austria, P. A., Kochhar, D., Tijerina, L., Biever, W., Diptiman, T., Hogsett, J., & Kiger, S. (2006). *Driver Workload Metrics Task 2 Final Report: (729342011-001)* [Data set]. American Psychological Association. https://doi.org/10.1037/e729342011-001

Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Davis, J., & Rodgers, J. (n.d.). *SCAPE: Shape Completion and Animation of People*.

Ann, O. C., & Theng, L. B. (2014). *Human Activity Recognition: A Review*.

Azhar, S., Kim, J., & Salman, A. (2018). *IMPLEMENTING VIRTUAL REALITY AND MIXED REALITY TECHNOLOGIES IN CONSTRUCTION EDUCATION: STUDENTS' PERCEPTIONS AND LESSONS LEARNED*. 3720–3730. https://doi.org/10.21125/iceri.2018.0183

Ballan, L., Taneja, A., Gall, J., Van Gool, L., & Pollefeys, M. (2012). Motion Capture of Hands in Action Using Discriminative Salient Points. In A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, & C. Schmid (Eds.), *Computer Vision – ECCV 2012* (Vol. 7577, pp. 640–653). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-33783-3_46

Bangaru, S. S., Wang, C., & Aghazadeh, F. (2022). Automated and Continuous Fatigue Monitoring in Construction Workers Using Forearm EMG and IMU Wearable Sensors and Recurrent Neural Network. *Sensors*, *22*(24), 9729. https://doi.org/10.3390/s22249729

Bo Li, Chunhua Shen, Yuchao Dai, van den Hengel, A., & Mingyi He. (2015). Depth and surface normal estimation from monocular images using regression on deep features and hierarchical CRFs. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1119–1127. https://doi.org/10.1109/CVPR.2015.7298715

Boton, C. (2018). Supporting constructability analysis meetings with Immersive Virtual Reality-based collaborative BIM 4D simulation. *Automation in Construction*, *96*, 1–15. https://doi.org/10.1016/j.autcon.2018.08.020

Camplani, M., & Salgado, L. (2012). *Efficient spatio-temporal hole filling strategy for Kinect depth maps* (A. M. Baskurt & R. Sitnik, Eds.; p. 82900E). https://doi.org/10.1117/12.911909

Caputo, F., Greco, A., D'Amato, E., Notaro, I., & Spada, S. (2019). IMU-Based Motion Capture Wearable System for Ergonomic Assessment in Industrial Environment. In T. Z. Ahram (Ed.), *Advances in Human Factors in Wearable Technologies and Game Design* (Vol. 795, pp. 215–225). Springer International Publishing. https://doi.org/10.1007/978-3-319-94619-1_21

Chen, C., Jafari, R., & Kehtarnavaz, N. (2015). Improving Human Action Recognition Using Fusion of Depth Camera and Inertial Sensors. *IEEE Transactions on Human-Machine Systems*, *45*(1), 51–61. https://doi.org/10.1109/THMS.2014.2362520

Chen, T., Yabuki, N., & Fukuda, T. (2020, October 14). *An Integrated Sensor Network Method for Safety Management of Construction Workers*. 37th International Symposium on Automation and Robotics in Construction, Kitakyushu, Japan. https://doi.org/10.22260/ISARC2020/0118

Chen, T., Yabuki, N., & Fukuda, T. (2021). *An Active Early Warning System for Heavy Construction Vehicle Drivers based on Mixed Reality*.

Cho, H.-J., & Tseng, M.-T. (2013). A support vector machine approach to CMOS-based radar signal processing for vehicle classification and speed estimation. *Mathematical and Computer Modelling*, *58*(1–2), 438–448. https://doi.org/10.1016/j.mcm.2012.11.003

Christian, M. S., Bradley, J. C., Wallace, J. C., & Burke, M. J. (2009). Workplace safety: A meta-analysis of the roles of person and situation factors. *Journal of Applied Psychology*, *94*(5), 1103–1127. https://doi.org/10.1037/a0016172

Cismas, A., Ioana, M., Vlad, C., & Casu, G. (2017). Crash Detection Using IMU Sensors. *2017 21st International Conference on Control Systems and Computer Science (CSCS)*, 672–676. https://doi.org/10.1109/CSCS.2017.103

Dai, F., Olorunfemi, A., Peng, W., Cao, D., & Luo, X. (2021). Can mixed reality enhance safety communication on construction sites? An industry perspective. *Safety Science*, *133*, 105009. https://doi.org/10.1016/j.ssci.2020.105009

Dehzangi, O., & Sahu, V. (2018). IMU-Based Robust Human Activity Recognition using Feature Analysis, Extraction, and Reduction. *2018 24th International Conference on Pattern Recognition (ICPR)*, 1402–1407. https://doi.org/10.1109/ICPR.2018.8546311

Dehzangi, O., Taherisadr, M., & ChangalVala, R. (2017). IMU-Based Gait Recognition Using Convolutional Neural Networks and Multi-Sensor Fusion. *Sensors*, *17*(12), 2735. https://doi.org/10.3390/s17122735

Eigen, D., & Fergus, R. (2015). *Predicting Depth, Surface Normals and Semantic Labels with a Common Multi-Scale Convolutional Architecture* (arXiv:1411.4734). arXiv. http://arxiv.org/abs/1411.4734

Elhayek, A., de Aguiar, E., Jain, A., Thompson, J., Pishchulin, L., Andriluka, M., Bregler, C., Schiele, B., & Theobalt, C. (2017). MARCOnI—ConvNet-Based MARker-Less Motion Capture in Outdoor and Indoor Scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(3), 501–514. https://doi.org/10.1109/TPAMI.2016.2557779

Elrawi, O. M. (2017). The Use of Mixed-Realities Techniques for the Representation of Islamic Cultural Heritage. *2017 International Conference on Machine Vision and Information Technology (CMVIT)*, 58–63. https://doi.org/10.1109/CMVIT.2017.16

Eskandarian, A., Sayed, R., Delaigue, P., Mortazavi, A., & Blum, J. (2008). *ADVANCED DRIVER FATIGUE RESEARCH*. *FMCSA-RRR-07-001*. https://doi.org/10.21949/1502667

ETS 2. (n.d.). Euro Truck Simulator 2. Retrieved April 19, 2023, from https://store.steampowered.com/app/227300/Euro_Truck_Simulator_2/

Filev, D., Lu, J., Prakah-Asante, K., & Tseng, F. (2009). Real-time driving behavior identification based on driver-in-the-loop vehicle dynamics and control. *2009 IEEE International Conference on Systems, Man and Cybernetics*, 2020–2025. https://doi.org/10.1109/ICSMC.2009.5346735

Google Glass. (n.d.). Google Glass. Retrieved April 19, 2023, from https://www.google.com/glass/start/

Guo, X., & Dai, Y. (2018). *Occluded Joints Recovery in 3D Human Pose Estimation based on Distance Matrix* (arXiv:1807.11147). arXiv. http://arxiv.org/abs/1807.11147

Haslam, R. A., Hide, S. A., Gibb, A. G. F., Gyi, D. E., Pavitt, T., Atkinson, S., & Duff, A. R. (2005). Contributing factors in construction accidents. *Applied Ergonomics*, *36*(4), 401–415. https://doi.org/10.1016/j.apergo.2004.12.002

Hayhoe, M. M. (2004). Advances in Relating Eye Movements and Cognition. *Infancy*, *6*(2), 267–274. https://doi.org/10.1207/s15327078in0602_7

Hirota, K., & Murakami, T. (2016). IMU Sensor based Human Motion Detection and Its Application to Braking Control of Electric Wheeled Walker for Fall-prevention. *IEEJ Journal of Industry Applications*, *5*(4), 347–354. https://doi.org/10.1541/ieejjia.5.347

HoloLens. (n.d.). HoloLens 2—Overview, Features, and Specs |. Retrieved April 19, 2023, from https://www.microsoft.com/en-us/hololens/hardware

Huang, Y., Bogo, F., Lassner, C., Kanazawa, A., Gehler, P. V., Romero, J., Akhter, I., & Black, M. J. (2017). Towards Accurate Marker-Less Human Shape and Pose Estimation over Time. *2017 International Conference on 3D Vision (3DV)*, 421–430. https://doi.org/10.1109/3DV.2017.00055

Intel. (n.d.). ® Distribution of OpenVINO™ Toolkit. Retrieved April 19, 2023, from https://www.intel.com/content/www/us/en/developer/tools/openvino-toolkit/overview.html

Jeelani, I., Han, K., & Albert, A. (2017). Development of Immersive Personalized Training Environment for Construction Workers. *Computing in Civil Engineering 2017*, 407–415. https://doi.org/10.1061/9780784480830.050

Kensaibou. (n.d.). Occurrence of Labor Disaster in Construction. Retrieved April 19, 2023, from https://www.kensaibou.or.jp/safe_tech/statistics/occupational_accidents.html

Khan, M. Q., & Lee, S. (2019). A Comprehensive Survey of Driving Monitoring and Assistance Systems. *Sensors*, *19*(11), 2574. https://doi.org/10.3390/s19112574

Khurshid, K., Danish, A., Salim, M. U., Bayram, M., Ozbakkaloglu, T., & Mosaberpanah, M. A. (2023). An In-Depth Survey Demystifying the Internet of Things (IoT) in the Construction Industry: Unfolding New Dimensions. *Sustainability*, *15*(2), 1275. https://doi.org/10.3390/su15021275

Kim, J., Hong, S., Baek, J., Kim, E., & Lee, H. (2012). *Autonomous vehicle detection system using visible and infrared camera*. 630–634.

Kim, K., Kim, H., & Kim, H. (2017). Image-based construction hazard avoidance system using augmented reality in wearable device. *Automation in Construction*, *83*, 390–403. https://doi.org/10.1016/j.autcon.2017.06.014

Kim, S.-N., & Lee, H. (2022). Capturing reality: Validation of omnidirectional video-based immersive virtual reality as a streetscape quality auditing method. *Landscape and Urban Planning*, *218*, 104290. https://doi.org/10/gpj5w5

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). *The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data: (729262011-001)* [Data set]. American Psychological Association. https://doi.org/10.1037/e729262011-001

Kun, A. L., Meulen, H. van der, & Janssen, C. P. (2018). Calling while Driving Using Augmented Reality: Blessing or Curse? *Presence: Teleoperators and Virtual Environments*, *27*(1), 1–14. https://doi.org/10.1162/pres_a_00316

Lai, K., Bo, L., Ren, X., & Fox, D. (2011). *A Large-Scale Hierarchical Multi-View RGB-D Object Dataset*.

Li, G., Eben Li, S., & Cheng, B. (2015). Field operational test of advanced driver assistance systems in typical Chinese road conditions: The influence of driver gender, age and aggression. *International Journal of Automotive Technology*, *16*(5), 739–750. https://doi.org/10.1007/s12239-015-0075-5

Li, H., Chan, G., & Skitmore, M. (2012). Multiuser Virtual Safety Training System for Tower Crane Dismantlement. *Journal of Computing in Civil Engineering*, *26*(5), 638–647. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000170

Li, H., Lu, M., Hsu, S.-C., Gray, M., & Huang, T. (2015). Proactive behavior-based safety management for construction safety improvement. *Safety Science*, *75*, 107–117. https://doi.org/10.1016/j.ssci.2015.01.013

Li, Z., Chen, L., Peng, J., & Wu, Y. (2017). Automatic Detection of Driver Fatigue Using Driving Operation Information for Transportation Safety. *Sensors*, *17*(6), 1212. https://doi.org/10.3390/s17061212

Liang, Y., & Lee, J. D. (2010). Combining cognitive and visual distraction: Less than the sum of its parts. *Accident Analysis & Prevention*, *42*(3), 881–890. https://doi.org/10.1016/j.aap.2009.05.001

Liao, L., Su, L., & Xia, S. (2017). Individual 3D Model Estimation for Realtime Human Motion Capture. *2017 International Conference on Virtual Reality and Visualization (ICVRV)*, 235–240. https://doi.org/10.1109/ICVRV.2017.00055

Liao, Y., Li, G., Li, S. E., Cheng, B., & Green, P. (2018). Understanding Driver Response Patterns to Mental Workload Increase in Typical Driving Scenarios. *IEEE Access*, *6*, 35890–35900. https://doi.org/10.1109/ACCESS.2018.2851309

Liu, F., Chunhua Shen, & Guosheng Lin. (2015). Deep convolutional neural fields for depth estimation from a single image. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5162–5170. https://doi.org/10.1109/CVPR.2015.7299152

Logitech. (n.d.). G923 TRUEFORCE Sim Racing Wheel for Xbox, Playstation and PC. Retrieved April 19, 2023, from https://www.logitechg.com/en-us/products/driving/g923-trueforce-sim-racing-wheel.html

Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., & Black, M. J. (2015). SMPL: A skinned multi-person linear model. *ACM Transactions on Graphics*, *34*(6), 1–16. https://doi.org/10.1145/2816795.2818013

Lu, Y., Li, Y., Skibniewski, M., Wu, Z., Wang, R., & Le, Y. (2015). Information and Communication Technology Applications in Architecture, Engineering, and Construction Organizations: A 15-Year Review. *Journal of Management in Engineering*, *31*(1), A4014010. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000319

Lun, R., & Zhao, W. (2015). A Survey of Applications and Human Motion Recognition with Microsoft Kinect. *International Journal of Pattern Recognition and Artificial Intelligence*, *29*(05), 1555008. https://doi.org/10.1142/S0218001415550083

Malekitabar, H., Ardeshir, A., Sebt, M. H., & Stouffs, R. (2016). Construction safety risk drivers: A BIM approach. *Safety Science*, *82*, 445–455. https://doi.org/10.1016/j.ssci.2015.11.002

Mandal, B., Li, L., Wang, G. S., & Lin, J. (2017). Towards Detection of Bus Driver Fatigue Based on Robust Visual Analysis of Eye State. *IEEE Transactions on Intelligent Transportation Systems*, *18*(3), 545–557. https://doi.org/10.1109/TITS.2016.2582900

Matyunin, S., Vatolin, D., Berdnikov, Y., & Smirnov, M. (2011). Temporal filtering for depth maps generated by Kinect depth camera. *2011 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, 1–4. https://doi.org/10.1109/3DTV.2011.5877202

Mekruksavanich, S., & Jitpattanakul, A. (2023). Automatic Recognition of Construction Worker Activities Using Deep Learning Approaches and

Wearable Inertial Sensors. *Intelligent Automation & Soft Computing*, *36*(2), 2111–2128. https://doi.org/10.32604/iasc.2023.033542

Miao, D., Fu, J., Lu, Y., Li, S., & Chen, C. W. (2012). Texture-assisted Kinect depth inpainting. *2012 IEEE International Symposium on Circuits and Systems*, 604–607. https://doi.org/10.1109/ISCAS.2012.6272103

Mixed Reality. (n.d.). What Is Mixed Reality? Retrieved April 19, 2023, from https://learn.microsoft.com/en-us/windows/mixed-reality/discover/mixed-reality

Miyaji, M., Kawanaka, H., & Oguri, K. (2009). Driver's cognitive distraction detection using physiological features by the adaboost. *2009 12th International IEEE Conference on Intelligent Transportation Systems*, 1–6. https://doi.org/10.1109/ITSC.2009.5309881

Mizumachi, M., Kaminuma, A., Ono, N., & Ando, S. (2014). Robust Sensing of Approaching Vehicles Relying on Acoustic Cues. *Sensors*, *14*(6), 9546–9561. https://doi.org/10.3390/s140609546

MRTK. (n.d.). GitHub - Microsoft/MixedRealityToolkit-Unity: Mixed Reality Toolkit (MRTK) Provides a Set of Components and Features to Accelerate Cross-Platform MR App Development in Unity. Retrieved April 19, 2023, from https://github.com/microsoft/MixedRealityToolkit-Unity

Nashashibi, F., & Bargeton, A. (2008). Laser-based vehicles tracking and classification using occlusion reasoning and confidence estimation. *2008 IEEE Intelligent Vehicles Symposium*, 847–852. https://doi.org/10.1109/IVS.2008.4621244

Ogunseiju, O. R., Gonsalves, N., Akanmu, A. A., Bairaktarova, D., Bowman, D. A., & Jazizadeh, F. (2022). Mixed reality environment for learning sensing technology applications in Construction: A usability study. *Advanced Engineering Informatics*, *53*, 101637. https://doi.org/10.1016/j.aei.2022.101637

Olorunfemi, A., Dai, F., Tang, L., & Yoon, Y. (2018). Three-Dimensional Visual and Collaborative Environment for Jobsite Risk Communication. *Construction Research Congress 2018*, 345–355. https://doi.org/10.1061/9780784481288.034

PoseNet. (n.d.). /Posenet at Master · Tensorflow/Tfjs-Models · GitHub. Retrieved April 19, 2023, from https://github.com/tensorflow/tfjs-models/tree/master/posenet

Preece, S. J., Goulermas, J. Y., Kenney, L. P. J., & Howard, D. (2009). A Comparison of Feature Extraction Methods for the Classification of Dynamic Activities From Accelerometer Data. *IEEE Transactions on Biomedical Engineering*, *56*(3), 871–879. https://doi.org/10.1109/TBME.2008.2006190

Preece, S. J., Goulermas, J. Y., Kenney, L. P. J., Howard, D., Meijer, K., & Crompton, R. (2009). Activity identification using body-mounted sensors—A review of classification techniques. *Physiological Measurement*, *30*(4), R1–R33. https://doi.org/10.1088/0967-3334/30/4/R01

Rabbani, M., Mia, Md. J., Khan, T., & Zarif, M. I. I. (2020). A Survey on RealSense: In context of Research and Application. *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 1–6. https://doi.org/10.1109/ICCCNT49239.2020.9225558

Raspberry Pi. (n.d.). Buy a Raspberry Pi 4 Model B –. Retrieved April 19, 2023, from https://www.raspberrypi.com/products/raspberry-pi-4-model-b/

Rhodin, H., Robertini, N., Richardt, C., Seidel, H.-P., & Theobalt, C. (2015). A Versatile Scene Model with Differentiable Visibility Applied to Generative Pose Estimation. *2015 IEEE International Conference on Computer Vision (ICCV)*, 765–773. https://doi.org/10.1109/ICCV.2015.94

Roetenberg, D., Luinge, H., & Slycke, P. (2013). *Xsens MVN: Full 6DOF Human Motion Tracking Using Miniature Inertial Sensors*.

Rogez, G., & Schmid, C. (2016). *MoCap-guided Data Augmentation for 3D Pose Estimation in the Wild* (arXiv:1607.02046). arXiv. http://arxiv.org/abs/1607.02046

Sarafianos, N., Boteanu, B., Ionescu, B., & Kakadiaris, I. A. (2016). 3D Human pose estimation: A review of the literature and analysis of covariates. *Computer Vision and Image Understanding*, *152*, 1–20. https://doi.org/10.1016/j.cviu.2016.09.002

Shin, D.-P., Gwak, H.-S., & Lee, D.-E. (2015). Modeling the predictors of safety behavior in construction workers. *International Journal of Occupational Safety and Ergonomics*, *21*(3), 298–311. https://doi.org/10.1080/10803548.2015.1085164

Shohet, I. M., Wei, H.-H., Skibniewski, M. J., Tak, B., & Revivi, M. (2019). Integrated Communication, Control, and Command of Construction Safety and Quality. *Journal of Construction Engineering and Management*, *145*(9), 04019051. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001679

Statista. (n.d.). Japan: Fatality Number in the Construction Industry 2021. Retrieved April 19, 2023, from https://www.statista.com/statistics/1274117/japan-fatality-number-accident-construction-industry/

Steven M. Lee. (n.d.). Roadside Construction Risks and Safety Tips. Retrieved April 19, 2023, from https://www.attorneystevelee.com/our-library/roadside-construction-risks-and-safety-tips/

Telea, A. (2004). An Image Inpainting Technique Based on the Fast Marching Method. *Journal of Graphics Tools*, *9*(1), 23–34. https://doi.org/10.1080/10867651.2004.10487596

TensorFlow Lite. (n.d.). | ML for Mobile and Edge Devices. Retrieved April 19, 2023, from https://www.tensorflow.org/lite

Varol, G., Romero, J., Martin, X., Mahmood, N., Black, M. J., Laptev, I., & Schmid, C. (2017). Learning from Synthetic Humans. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4627–4635. https://doi.org/10.1109/CVPR.2017.492

von Marcard, T., Rosenhahn, B., Black, M. J., & Pons-Moll, G. (2017). Sparse Inertial Poser: Automatic 3D Human Pose Estimation from Sparse IMUs. *Computer Graphics Forum*, *36*(2), 349–360. https://doi.org/10.1111/cgf.13131

Wang, X., & Dunston, P. S. (2013). Tangible mixed reality for remote design review: A study understanding user perception and acceptance. *Visualization in Engineering*, *1*(1), 8. https://doi.org/10.1186/2213-7459-1-8

Weiss, A., Hirshberg, D., & Black, M. J. (2011). Home 3D body scans from noisy image and range data. *2011 International Conference on Computer Vision*, 1951–1958. https://doi.org/10.1109/ICCV.2011.6126465

Wu, S., Hou, L., Zhang, G. (Kevin), & Chen, H. (2022). Real-time mixed reality-based visual warning for construction workforce safety. *Automation in Construction*, *139*, 104252. https://doi.org/10.1016/j.autcon.2022.104252

Xu, J., & Lu, W. (2018). Smart Construction from Head to Toe: A Closed-Loop Lifecycle Management System Based on IoT. *Construction Research Congress 2018*, 157–168. https://doi.org/10.1061/9780784481264.016

Yin, J., Zhu, D., Shi, M., & Wang, Z. (2019). Depth Maps Restoration for Human Using RealSense. *IEEE Access*, *7*, 112544–112553. https://doi.org/10.1109/ACCESS.2019.2934863

Zhang, Z., Schwing, A. G., Fidler, S., & Urtasun, R. (2015). Monocular Object Instance Segmentation and Depth Ordering with CNNs. *2015 IEEE International Conference on Computer Vision (ICCV)*, 2614–2622. https://doi.org/10.1109/ICCV.2015.300

Zhou, X., Huang, Q., Sun, X., Xue, X., & Wei, Y. (2017). Towards 3D Human Pose Estimation in the Wild: A Weakly-Supervised Approach. *2017 IEEE International Conference on Computer Vision (ICCV)*, 398–407. https://doi.org/10.1109/ICCV.2017.51

Appendix

Program codes:

Scripts based on Unity 3d introduced in Chapter 5, such as detecting the three-dimensional coordinates of the hand, and detecting the nodding and shaking of the head are introduced below.

HandDetector:

```
using System.Collections;
using System.Collections.Generic;
using UnityEngine;
using Microsoft.MixedReality.Toolkit.Utilities;
using Microsoft.MixedReality.Toolkit.Input;
/// <summary>
/// This class detects the HoloLens 2's HandTracking joints.
/// </summary>
public class HandDetector : MonoBehaviour
{
    [SerializeField, HeaderAttribute("DetectTargetHand")]
    HandMode HandDetectMode;
    Handedness handednesstype;
    enum HandMode
    {
        RightHand,
        LeftHand,
        BothHand,
    }
    // Start is called before the first frame update
    void Start()
    {
        //DetectRighitHandWrist
        if ((int)HandDetectMode == 0)
        {
            handednesstype = Handedness.Right;
        }
        //DetectLeftHandWrist
        if ((int)HandDetectMode == 1)
        {
            handednesstype = Handedness.Left;
        }
        //DetectBothHandWrist
        if ((int)HandDetectMode == 2)
        {
            handednesstype = Handedness.Both;
        }
        Debug.Log(handednesstype);
    }
    // Update is called once per frame
    void Update()
    {
        //DetectRighitHandWrist
        if (HandJointUtils.TryGetJointPose(TrackedHandJoint.Wrist, handednesstype, out
MixedRealityPose pose))
        {
            Debug.Log("Detect");
            Debug.Log(pose);
        }
    }
}
```

HeadGestureRecognizer:

```
using System.Collections.Generic;
using UnityEngine;
using System.Linq;
```

```
using System;
/// <summary>
/// Judgment class for head gesture
/// </summary>
public class HeadGestureRecognizer : MonoBehaviour
{
    /// <summary>
    /// Head rotation gesture type
    /// </summary>
    public enum HeadRotateGesture
    {
        /// <summary>
        /// None (default)
        /// </summary>
        Nothing = 0,
        /// <summary>
        /// nod
        /// </summary>
        Nod = 1,
        /// <summary>
        /// swinging head
        /// </summary>
        Shake = 2,
        /// <summary>
        /// tilting head
        /// </summary>
        Tilt = 3,
    }
    /// <summary>
    /// Rotation Pose Sampling Data Type
    /// </summary>
    public struct PoseSample
    {
        // Time stamp
        public readonly float Timestamp;
        // Direction of rotation
        public Quaternion Orientation;
        // euler angles
        public Vector3 EulerAngles;
        public PoseSample(float timestamp, Quaternion orientation)
        {
            Timestamp = timestamp;
            Orientation = orientation;
            EulerAngles = orientation.eulerAngles;
            EulerAngles.x = WrapDegree(EulerAngles.x);
            EulerAngles.y = WrapDegree(EulerAngles.y);
            EulerAngles.z = WrapDegree(EulerAngles.z);
        }
        /// <summary>
        /// Convert Euler angles to the range 180 degrees to -180 degrees
        /// </summary>
        public float WrapDegree(float degree)
        {
            if (degree > 180f)
            {
                return degree - 360f;
            }
            return degree;
        }
    }
    /// <summary>
    /// Head rotation gesture type
    /// </summary>
    [SerializeField, Tooltip("Head rotation gesture type")]
    private HeadRotateGesture p_HeadRotateGesture = HeadRotateGesture.Nothing;
    /// <summary>
    /// Normal event
    /// </summary>
    public Action EventNothing;
```

```
/// <summary>
/// nodding event
/// </summary>
public Action EventNod;
/// <summary>
/// bobble event
/// </summary>
public Action EventShake;
/// <summary>
/// tilt event
/// </summary>
public Action EventTilt;
/// <summary>
/// Rotation Pose Cues
/// </summary>
public readonly Queue<PoseSample> PoseSamples = new Queue<PoseSample>();
/// <summary>
/// Event interval time (seconds)
/// </summary>
[SerializeField]
private float recognitionInterval = 0.5f;
/// <summary>
/// Last Gesture Occurrence Time
/// </summary>
private float prevGestureTime;
/// <summary>
/// Gesture execution flag
/// </summary>
private bool p_GesturedFlg;
/// <summary>
/// Periodic processing
/// </summary>
void Update()
{
    // Get the current head local rotation
    var orientation = Camera.main.transform.localRotation;
    // Queue rotation information with timestamp
    PoseSamples.Enqueue(new PoseSample(Time.time, orientation));
    if (PoseSamples.Count >= 120)
    {
        //Store up to 120 queues
        PoseSamples.Dequeue();
    }
    // Between the last gesture event and the interval time
    // Do not perform new gesture judgment
    if (!(prevGestureTime < Time.time - recognitionInterval)) return;
    // Turn off gesture judgment flag
    p_GesturedFlg = false;
    // Detect nodding gesture
    if (!p_GesturedFlg)
    {
        if (IsRecognizeNod())
        {
            p_HeadRotateGesture = HeadRotateGesture.Nod;
            // run the event
            EventNod?.Invoke();
            // Record the judgment time of the gesture and turn on the flag
            prevGestureTime = Time.time;
            p_GesturedFlg = true;
        }
    }
    // Judging the shaking gesture
    if (!p_GesturedFlg)
    {
        if (IsRecognizeShake())
        {
            p_HeadRotateGesture = HeadRotateGesture.Shake;
            // run the event
            EventShake?.Invoke();
```

```
                // Record the judgment time of the gesture and turn on the flag
                prevGestureTime = Time.time;
                p_GesturedFlg = true;
            }
        }
        // Judgment of tilting gesture
        if (!p_GesturedFlg)
        {
            if (IsRecognizeTilt())
            {
                p_HeadRotateGesture = HeadRotateGesture.Tilt;
                // run the event
                EventTilt?.Invoke();
                // Record the judgment time of the gesture and turn on the flag
                prevGestureTime = Time.time;
                p_GesturedFlg = true;
            }
        }
        //Did all gesture tests fail?
        if (!p_GesturedFlg)
        {
            // Judge as no gesture
            p_HeadRotateGesture = HeadRotateGesture.Nothing;
            // run the event
            EventNothing?.Invoke();
        }
    }
    /// <summary>
    /// Get rotation pose for specified time range
    /// </summary>
    IEnumerable<PoseSample> Range(float startTime, float endTime) =>
        PoseSamples.Where(sample =>
            sample.Timestamp < Time.time - startTime &&
            sample.Timestamp >= Time.time - endTime);
    /// <summary>
    /// Nodding judgment check
    /// </summary>
    private bool IsRecognizeNod()
    {
        bool isNod = false;
        try
        {
            // Get the average vertical rotation between 0.4 seconds and 0.2 seconds
ago
            var averagePitch = Range(0.2f, 0.4f).Average(sample => sample.EulerAngles.x);
            // Get the maximum value of vertical rotation (positive direction: downward
rotation) from 0.2 seconds ago to the present
            var maxPitch = Range(0.01f, 0.2f).Max(sample => sample.EulerAngles.x);
            //Get the latest vertical rotation angle
            var pitch = PoseSamples.Last().EulerAngles.x;
            // The maximum downward rotation angle is 5 degrees or more than the average
rotation angle.
            // And whether the latest rotation angle is 2.5 degrees or more back from
the maximum downward rotation angle
            if (!(maxPitch - averagePitch > 5.0f)
                || !(maxPitch - pitch > 2.5f)) return isNod;
            Debug.Log("Nod last : " + pitch + ", average : " + averagePitch
                + ", max : " + maxPitch);
            // Determine that a nod has occurred
            isNod = true;
        }
        catch (InvalidOperationException)
        {
            // Range contains no entry
        }
        return isNod;
    }
    private bool IsRecognizeShake()
    {
```

```
        bool isShake = false;
        try
        {
            // Get the average sideways rotation between 0.4s and 0.2s ago
            var averageYaw = Range(0.2f, 0.4f).Average(sample => sample.EulerAngles.y);
            // Get the maximum horizontal rotation value (positive direction: right
rotation) from 0.2 seconds ago to the present
            var maxYaw = Range(0.01f, 0.2f).Max(sample => sample.EulerAngles.y);
            // Get the minimum horizontal rotation value (negative direction: left
rotation) from 0.2 seconds ago to the present
            var minYaw = Range(0.01f, 0.2f).Min(sample => sample.EulerAngles.y);
            // Get the latest horizontal rotation angle
            var yaw = PoseSamples.Last().EulerAngles.y;
            // If the maximum rotation angle is not more than 10 degrees greater than
the average rotation angle, it is not swinging.
            if (!(maxYaw - averageYaw > 5.0f) ||
                !(averageYaw - minYaw > 5.0f)) return isShake;
            Debug.Log("Shake last : " + yaw + ", average : " + averageYaw
                + ", max : " + maxYaw + ", min : " + minYaw);
            // Determine that a swing has occurred
            isShake = true;
        }
        catch (InvalidOperationException)
        {
            // Range contains no entry
        }
        return isShake;
    }
    /// <summary>
    /// Tilt judgment check
    /// </summary>
    private bool IsRecognizeTilt()
    {
        bool isTilt = false;
        try
        {
            // Get the average frontal rotation from 0.4 seconds ago to now
            var averageTilt = Range(0.01f, 0.4f).Average(sample => sample.EulerAngles.z);
            // Get the angle of the latest frontal rotation
            var tilt = PoseSamples.Last().EulerAngles.z;
            // Is the average rotation angle in the front direction 20 degrees or more?
            if (!(averageTilt > 20.0f) &&
                !(averageTilt < -20.0f)) return isTilt;
            Debug.Log("Tilt last : " + tilt + ", average : " + averageTilt);
            // Determine that a head tilt has occurred
            isTilt = true;
        }
        catch (InvalidOperationException)
        {
            // Range contains no entry
        }
        return isTilt;
    }
}
```