ICCEPM 2020
The 8th International Conference on Construction Engineering and Project Management
Dec. 7-8, 2020, Hong Kong SAR

Joint Reasoning of Real-time Visual Risk Zone Identification and Numeric Checking for Construction Safety Management

Ahmed Khairadeen Ali[1]*, Numan Khan[1], Lee Do Yeop[1] , Chansik Park[1]*

[1] *1 School of Architecture and Building Science, Chung Ang University, Republic of Korea,* E-mail address:ahmedshingaly@gmail.com, *numanpe@gmail.com*, *doyeop@cau.ac.kr*, cpark@cau.ac.kr

Abstract: The recognition of the risk hazards is a vital step to effectively prevent accidents on a construction site. The advanced development in computer vision systems and the availability of the large visual database related to construction site made it possible to take quick action in the event of human error and disaster situations that may occur during management supervision. Therefore, it is necessary to analyze the risk factors that need to be managed at the construction site and review appropriate and effective technical methods for each risk factor. This research focuses on analyzing Occupational Safety and Health Agency (OSHA) related to risk zone identification rules that can be adopted by the image recognition technology and classify their risk factors depending on the effective technical method. Therefore, this research developed a pattern-oriented classification of OSHA rules that can employ a large scale of safety hazard recognition. This research uses joint reasoning of risk zone Identification and numeric input by utilizing a stereo camera integrated with an image detection algorithm such as (YOLOv3) and Pyramid Stereo Matching Network (PSMNet). The research result identifies risk zones and raises alarm if a target object enters this zone. It also determines numerical information of a target, which recognizes the length, spacing, and angle of the target. Applying image detection joint logic algorithms might leverage the speed and accuracy of hazard detection due to merging more than one factor to prevent accidents in the job site.

Keywords: Risk Zone Identification, Numeric Checking, Image Recognition, Safety in Construction, Depth Estimation

1. INTRODUCTION

Despite various efforts to reduce the number of accidents and fatality in construction sites, construction Safety accidents are occurring continuously. safety in construction remains a critical issue. In the up to date records of the Occupational Safety and Health Administration (OSHA), 169 fatality cases of workers struck by vehicles where registered. The number of fatalities, wounds, and close misses is that they present liabilities that can be avoided. Safe development requires care and arranging all through the undertaking life-cycle, from the structure, through development arranging, through development execution, and reaching out into tasks and records [1]. The conventional safety management in the construction industry is time consuming, costly, inefficient, and hard to control in big size projects [2], [3]. Therefore, OSHA needs to be deeply analyzed and checked if the current state of the art advancement in automatic rule-checking technologies can adopt these rules to leverage a safe environment in the construction site.

Nowadays there has been advancement in construction monitoring and rule automatic checking such as drone monitoring and simulation, equipment/material connectivity and tracking, robotics and automated technology, sensors, and reporting platforms, building information modeling. However, most of these technologies include limitations by handling a specific task at a time. As a result, Image Recognition Technology (IRT) is the most economically efficient, complex pattern recognition, visual-based risk recognition which is similar to the safety manager visual judgment process [4]–[6].

Image recognition-related technologies have the advantage of being able to systematically identify destructive behavior and unsafe environments without affecting the productivity of workers in the field, with little additional cost to expand separate devices after system algorithms are deployed [7], [8]. Besides, the safety manager's discovery of risk factors is crucially dependent on visual judgment, and the application of the proposed technology to replace human eyes is very high [9]. Concurrently, this technology is considered effective in helping or replacing some of the observer's tasks, since the site usability can be ensured even in his absents [10]–[12]. Recent researches proposed an automated inspection system implementing rule-based algorithms and analyzed models automatically to detect danger and tolerate preventive actions. The prototype was developed to automatically reflect falls prevention measures such as covers, and temporary rails installed to prevent workers from falling in wholes [13], [14] .

However, existing prior studies using visual recognition technology have been limited to studying the applicability of this technology and improving accuracy in several cases of the unsafe environment or risky behavior in the site, and for the proposed research system to be feasible in practical use, it is necessary to identify danger factors so that various hazards existing at the site can be detected simultaneously. Besides, this technology should be introduced in the optimal location at the Planning (PLAN) stage because it performs merely within the viewing line and generates vast amounts of data while transmitting and processing when used in the absolute space, depending on the location of the camera installation. Therefore, the state-of-the-art researches focusing on Rule checking safety detection and image detection technologies were reviewed and analyzed to determine risk zones around hazard objects and detect if any target (person) crossing it.

2. LITERATURE REVIEW

A thrust of scholars analyzed the risk factors that might occur during heavy equipment work and proposed a measure to recognize the danger by utilizing a stereo camera to non-electric-based technology. In the study, three major threat factors were identified: speed of equipment, access to perilous places, and proximity between two objects, and the study was conducted on loading, transporting, and unloading operations. This study developed an algorithm code based on C++ language and presented a way to judge threat factors on the visual base [15]. To attach Radio Frequency Identification (RFID) devices to heavy equipment to prevent workers and heavy equipment crashes such as excavators and cranes. By doing so, it was derived that it was sufficiently preventable in case of collision due to the failure of pre-inspection [16], [17]. Fang (2018) applied Faster-R-CNN to determine whether to wear a protective helmet for job site workers and randomly selected more than 100,000 construction worker image frames after filming at 25 job sites. The classification items were primarily divided into Weather, Illumination, Individual Posture, Visual range, and Occlusion, and 19 classifications were made according to the specific classification, which was applied to the Faster-R-CN algorithm to effectively recognize workers who are not in the use of helmets [19]–[22].

To leverage the construction safety management tasks, various technologies such as automatic design, sensor, and location tracking are being developed. However, the sensor, location-based technology applied to prior studies must be equipped with relevant devices in the worker's body or helmet, thereby reducing the efficiency of the work [23], [24]. Besides, devices are installed on a per-target basis, additional work is done to manage various components, and overspending is inevitable as the scope of control increases.

The presented study aims to progressively expand the possible target of visual technology for automatic rule checking by exploring OSHA rules and construction job site accident history and then applying image detection algorithm measures for safety management in the planning and construction phase. The real-time object detection YOLOv3 with its accuracy and speed can be integrated with the stereo algorithms to detect and measure depth simultaneously [15], [25]. The state-of-the-art depth estimation using CNN lacks the means to exploit feature information of the stereo pair of images. Therefore, this research reasonably used Pyramid Stereo Matching Network (PSMNet) to specialize in the pyramid pooling and typically utilize 3D CNN to regularize cost volume [26], [27]. The state of the art researches employing image recognition was properly classified under three groups: scene based risk identification, location-based risk identification, and action-based risk identification [28]. However, this classification didn't cover the variety of safety control on the job site. Therefore, this research developed a pattern-oriented classification of OSHA rules that can employ joint reasoning of zone identification and numeric checking in the construction job site by integrating YOLOv3 and PSMNet inside stereo depth camera.

3. METHODOLOGY

Numerical determination (length, spacing, angle)

Determine numeric information of a target, which includes numerical information of the length, spacing, and angle of the target and raise alarm when measurements are outside the acceptable threshold by comparison with the design literature or related statutes, and includes the installation interval and installation angle of the protective materials installed primarily for work during construction phases.

Risk zone identification

This type of risk assessment recognizes sets of the danger zone and checks whether workers enter or evacuate the determined zone. Examples include lifting work using cranes where the cargo traces path access zone should be controlled and prohibited to cross to prevent cargo passing over the workers' head as shown in Figure 1. Equally, it includes prohibiting the workers from entering the loading area to prevent struck or collision with the unloading machines such as forklifts.

The person-vehicle risk zone determination scenario was chosen (as highlighted in the yellow box) as a proof of concept to test whether it is possible to merge numeric determination and risk zone identification using one monitoring device. In this case, the numeric input will remain the minimum distance between the target which is the job site worker, and the hazardous object which represents the closest vehicle. The risk zone identification will be surrounding the vehicle boundaries to determine whether a target will enter or exit this zone. The numeric input and the risk zone diameter will depend on OSHA rules that are pre-defined and are used as input in the below algorithm as illustrated in Figure 2.

The methodological framework consists of five steps. The risk zone diameter and numeric input are referenced inside the kitti stereo data scene flow 2015 algorithms. In the second step, YOLOv3 algorithms run the stereo camera recording and detect the target (person) and hazard object (car). The third step runs the PSMNet algorithms to determine the distance between the camera-target (person) and camera- hazard object (car 1). In the fourth step, the research uses a mathematical triangulation equation to calculate the distance between the target (person) and the hazard object (car 1). In the final step, the algorithm determines the risk zone boundaries, distance to the target, and draw a red box if the target enters the risk zone as illustrated in Figure 2.

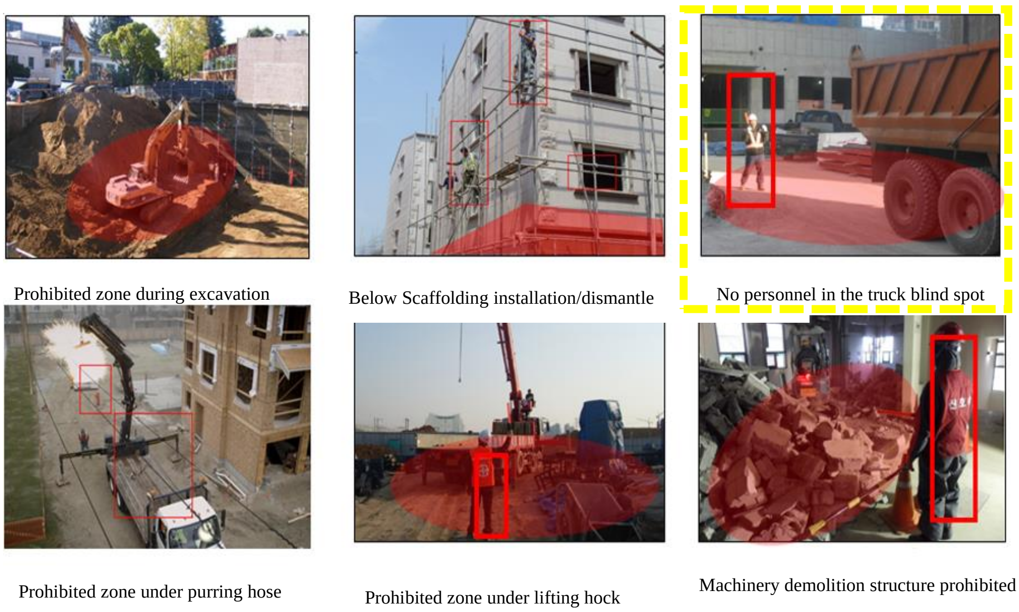

Figure 1. Different scenarios of determining prohibited zone during construction

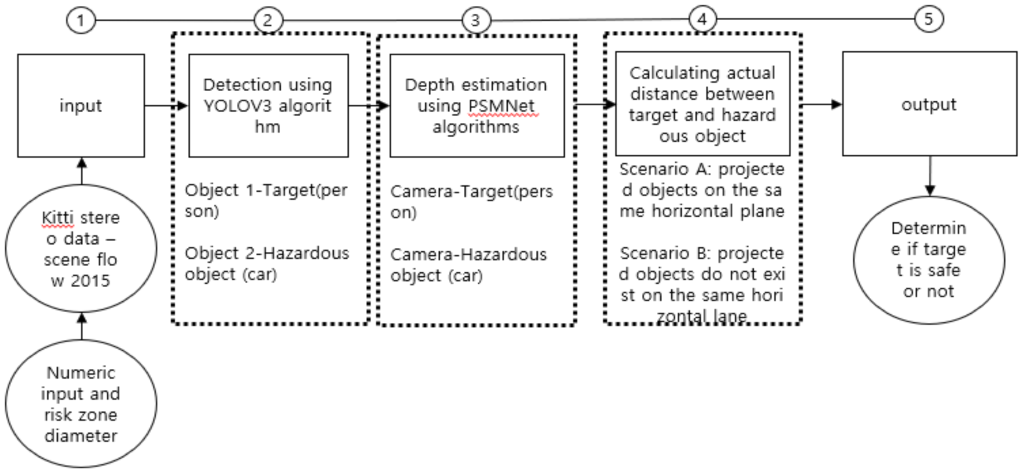

Figure 2. Research methodology framework

4. PROOF OF CONCEPT

4.1. Input

The numeric input and risk zone identification determined in the OSHA rules were pre-defined and adjusted to work for this specific example. The kitti stereo depth data – scene flow evaluation 2015 kit [29] was used as input data. This benchmark contains RGB, monochrome GPS, and laser scanner. The value of stereo confidence left-right consistency check of disparity of the pre-recorded data set was also used as input in the kitti flow evaluation. Its benchmark consists of 400 training and testing scenes. The pixel disparity is estimated to be less than 5% which is vital to determine an accurate distance calculation close to the actual measurements. The data were used to train PSM-Net and test its results shown in Figure 3.

Figure 3. RGB images from kitti depth data – scene flow evaluation 2015

4.2. Detection

There-recorded data from the kitti stereo– scene flow evaluation 2015 was then inputted into YOLO v3 to detect the visible objects in the scene as illustrated in the figure below. Yolo v3 is a detection algorithm suitable for use in construction sites where real-time detection is possible. The output of Yolo v3. Percentage means confidence in the model. The next line is the coordinate influenced by the model and it is represented in pixels. For example, car 1's bounding box corner coordinates are left top is (624,163), right top is (1056,163), left bottom is (624,330) and right bottom is (1056,330). The target (person) and the closest hazardous object (car 1) were detected and verified. Their bounding box corner x and y coordinate data were extracted from YOLOv3 and stored for the depth estimation using a different algorithm as illustrated in Figure 4.

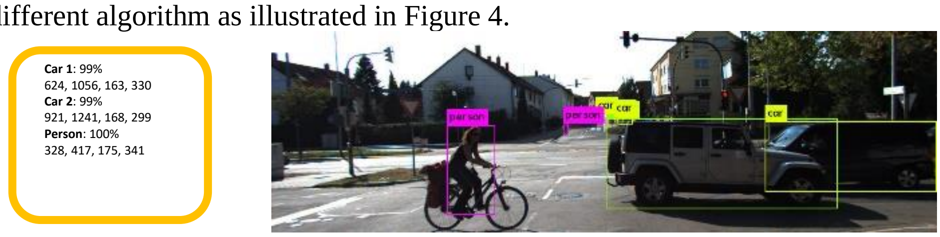

Figure 4. Image detection results in running YOLOv3 algorithms

4.3. Depth estimation

The same initial data are input into the deep learning network model (PSMNet). This algorithm extracted the feature values of the stereo/mono image and created a 2D feature map. Then it used 3D CNN to match cost computation. It finally predicted the disparity map. Each point in the disparity regression image represents the distance from the camera as shown in the below figure. The distance between camera-target (person) and camera-hazardous object (car) was extracted from the depth values. The final step is to Select the location-to-value of the detected Target and hazardous object in YOLOv3 to obtain the actual depth as illustrated in Figure 6. The center of the bounding box is considered the starting point to measure a straight line to the field of view center of the stereo camera when there is no overlap between objects Centers. However, Overlaid the center of the uncovered area will be the starting point to measure a start a linear measurement toward the field of view center of the stereo camera when the Object's Center Is Blocked.

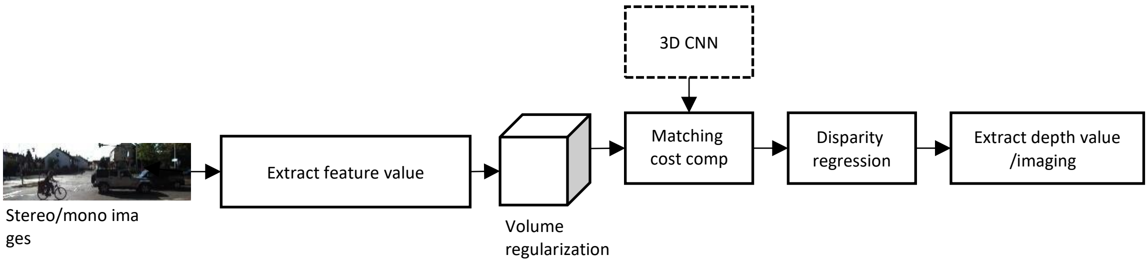

Figure 5. Deep learning model (PSMNet) working process

Figure 6. PSINet depth estimation regression map

4.4. Calculating the actual distance between the target object and the hazard object

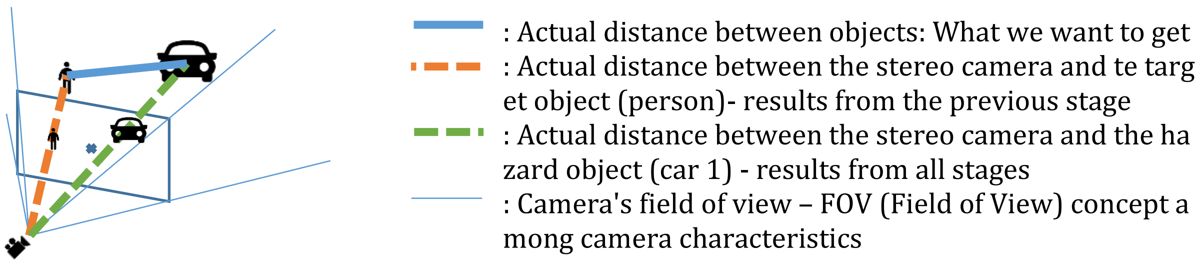

Figure 7. the relationship between the viewing angle of the stereo camera and the number of pixels

Scenario A: If the projected objects are on the same horizontal line (yellow line)

The green line in the middle of the below figure: Cross the center of the image in the direction in front of the stereo camera. As a result, the distance between the stereo camera and the target (person), the camera, and the hazardous object (car 1) can be calculated using the PSMNet output data. Then the actual distance between the target (person) and hazard object (car 1) can be found using the below formula as illustrated in Figure 8.

α: The angle between the camera's front vector and the camera to a person's vector

x: the number of pixels between a person and the center

width: number of lateral pixels in the image

tan (visual/2) = width/2/a

tan(α) = x/a

$$\alpha = \arctan\left(\frac{2x\tan\left(\frac{visual}{2}\right)}{width}\right)$$

β can be obtained using the same formula

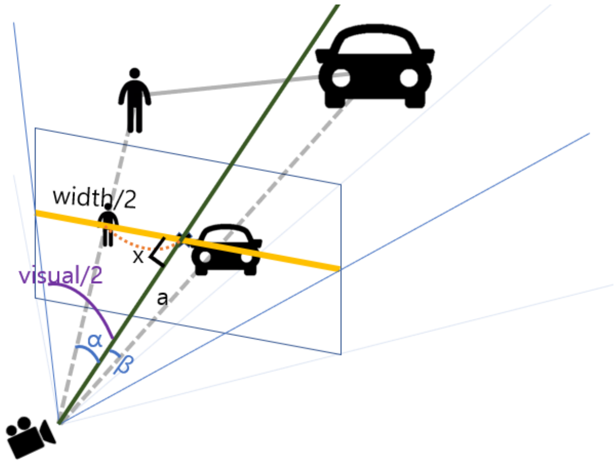

Figure 8. Step 4. Scenario A. projected objects on the same horizontal line

Scenario B. The projected objects do not exist on the same horizontal plane

If the target object (person) and the hazard object (car 1) do not exist on the same horizontal line, calculate the number of pixels between the original position and the moved position. In this case (α) can be generated by using the same value as the previous width FOV, where the dark green solid line is the distance * sin (α) from the stereo camera to the hazard object (car 1). The length of the yellow dotted line can be calculated by multiplying the length of the green dotted line by the cosine of (α) value. The length of the orange dotted line can be generated using the stereo camera depth calculation data. if the length of the yellow dotted line and the orange dotted line are known then draw a horizontal blue line between the target and the hazard object. As a result, the angle between the yellow dotted line and the orange dotted line can be calculated using the triangulation equation method shown in the previous example. Besides, the distance between the target (person) and the hazard object (car1) can also be generated using the previous triangulation equation as represented by the continuous blue line as shown in Figure 9.

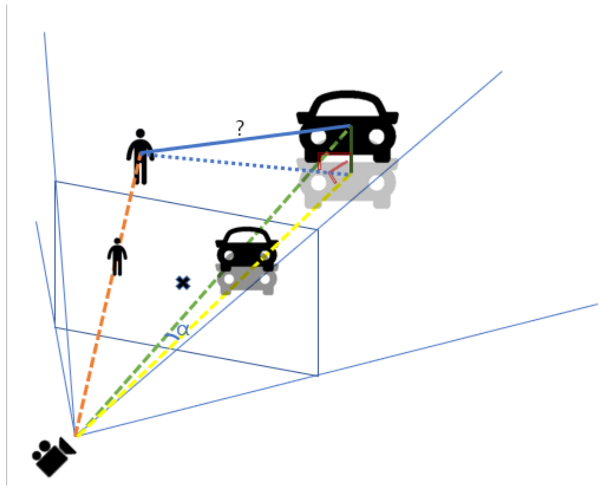

Figure 9. Step 4. Scenario B. projected objects, not on the same horizontal line

4.5. Output

In the illustrated scenario, If the distance between the target object and the hazardous object is closer than the specified distance determined in the OSHA rules or decided by the safety manager in the construction job site, the target object (person) is represented by a red box. Furthermore, an alarm should be sent to the target, hazardous object (car 1) operator and the Jobsite safety manager to eliminate the detected unsafe behavior as shown in the below figure.

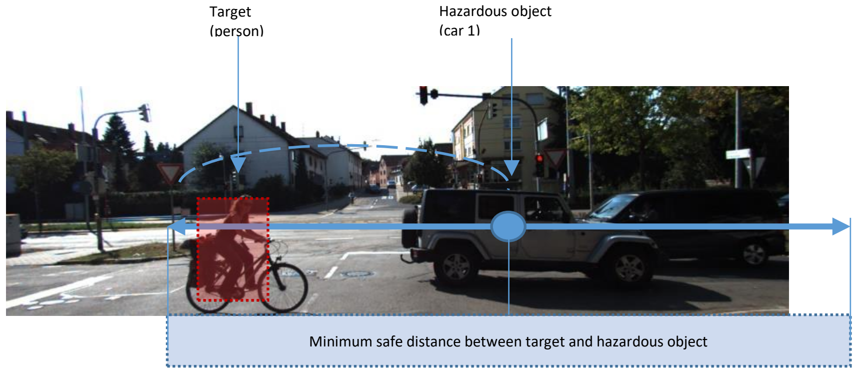

Figure 10. Illustration figure of the risk zone and target detecting red box

5. PROPOSED MONITORING ADVANCEMENT

The conventional numeric judgment is determined by using measurements after discovering facilities that are significantly in compliance with OSHA standards depending on the supervisor's experience and knowledge. If the allowable distance and allowable angle are exceeded, the nearby workers should seize work, reinstall the facility following the criteria then process work as illustrated in Figure 11.

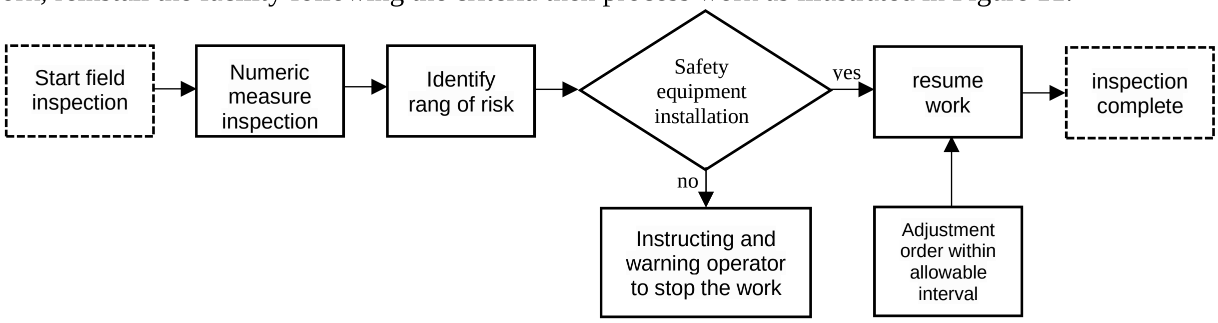

Figure 11. Conventional numeric inspection As-Is workflow

In contrast, when the proposed technology is applied on-site, it will be able to check and measure the protection, temporary kits installed and send warning automatically when they don't match the standards as presented in Figure 12.

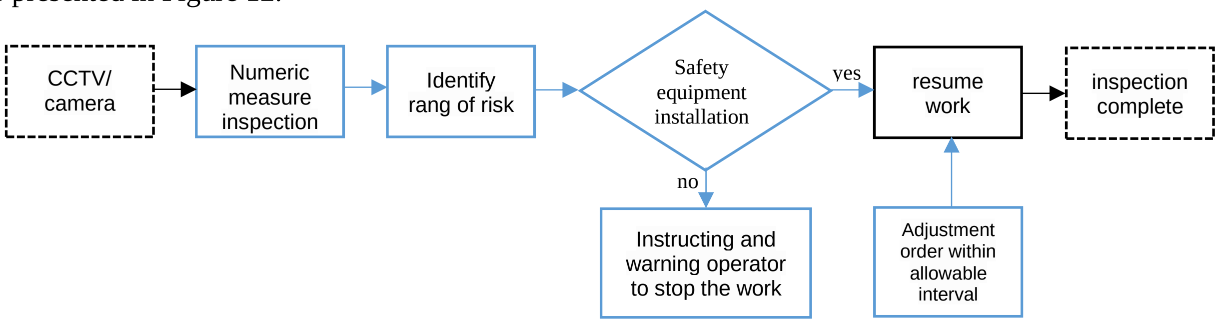

Figure 12. Numeric Determination to-be process

In the conventional risk zone identification workflow, the safety manager shall determine the place where access to the work should be prohibited and give direct instructions to the work team leader or guide/signal number, but it is not professional and often overlooked for the efficiency of the work as illustrated in Figure 13.

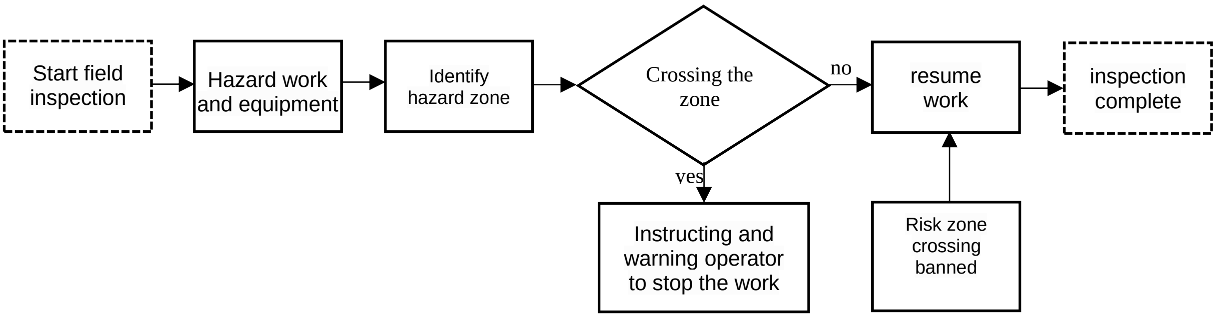

Figure 13. Risk zone As-Is workflow

The proposed methodology can reduce the workload by replacing some of the safety managers' tasks during the building process because of the constantly changing site conditions and the characteristics of the sites where various types of work occur simultaneously as illustrated in Figure 14.

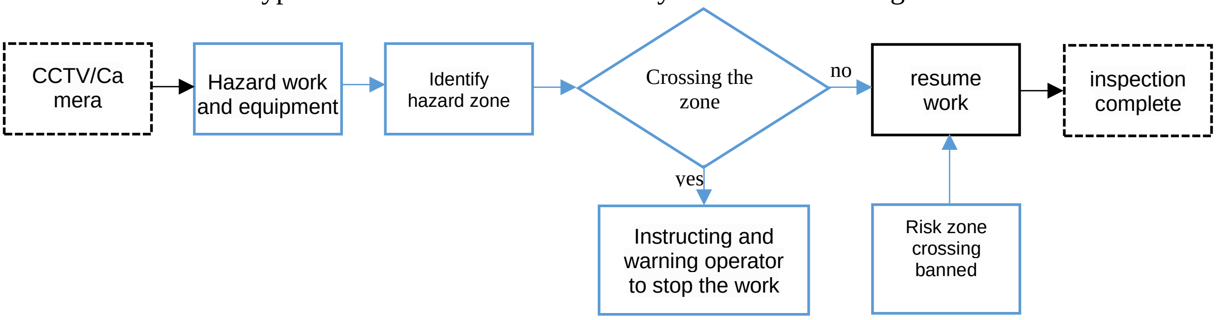

Figure 14. Risk zone identification To-Be workflow

5. DISCUSSION

The risk zone identification in the construction site in different scenarios shares similar characteristics such as occupying a circular space, preventing a target to enter or exit the zone, and having a fixed diameter. However, the numeric checking of this zone concerning a specific target (worker) might vary from one situation to another. For example, in the risk zone identification relationship between worker-vehicle is often horizontal. This research only covered one scenario which is worker-vehicle risk zone identification space and tested the algorithms for it. However, the relationship between worker-lifting hock is vertical which requires different numeric inputs. Therefore, the proposed algorithm should be modified for each activity independently. Also, for the risk zone identification and numeric checking to work, this research proposes integrating two different algorithms (YOLOv3 for image detection and PSMNet for depth calculation), which burden the monitoring and take a long time to process. It might be possible to merge both algorithms into one platform that detect objects and measure the distance between them simultaneously. This research only used the database prerecorded database of Kitti stereo depth– scene flow evaluation 2015 kit as a proof of concept. However, the construction Jobsite is a continuously changing environment with a variety of activities occurring at the same time, which makes the detection harder than the pre-recorded data that might add to the proposed methodology's challenges.

The process of recording the scene, importing it to the image detection algorithm, import the data of the detected object into PSMnet then finally calculate the distance and comparing it to OSHA rules is manual and consumes a lot of time. This research is pushing toward converting this process into a semi-automatic using only two platform one for detection and one for depth calculation. The rest steps are to be embedded within the two platform algorithms.

The proposed joint reasoning, risk detection, and numeric checking using image recognition technologies can replace or enhance the safety manager's performance in the construction job site. The conventional safety management takes a lot of time and lacks efficiency due to the requirement of physical existence at the Jobsite frequently. Therefore, the proposed method might be able to replace the safety manager's judgment of risk identification in the specific activities highlighted in the methodology section.

5. CONCLUSION AND FUTURE WORK

The integration between image detection YOLOv3 and PSMNet imbedded in the Stereo camera enabled the possibility to determine a risk zone identification around hazard object (e.g. Vehicle) and checks the numeric distance from the risk zone center of a target object (e.g. Worker) on the construction job site. Also, many researchers tried to reduce disasters in the construction site by applying image detection or depth estimation algorithms, but the limitation was fragmentary, and the focus of the research was on the applicability and accuracy of the technology. Recent scholars explored the possibility of reducing disasters in the construction job sites by applying image recognition technology, but the limitation was fragmentary, and the focus of the research was on the applicability and accuracy of the technology.

Therefore, to ensure the feasibility and applicability of the visual detection among the venture in the construction site, the risk factors are judged by the safety manager's visual and measurement tools. Furthermore, through the application of visual detection, the analysis process was automated to reduce the workload of safety managers who lacked manpower in sites. Due to the nature of the work of building, it is inevitable to rely on experience, along with visual judgment, and the inconvenience of the observer visiting the site exists.

In the future, test the same scenario using other available depth estimation algorithms such as the CSPN might give different valuable assets to risk zone identification in the construction job site. Also, the algorithm result can be tested using actual construction Jobsite PSMNnet physical data, photos instead of using prerecorded data available with the algorithm. Finally, PSMNet fine-tuning requires more learning data and ground truth to raise the accuracy of depth and distance prediction.

ACKNOWLEDGMENTS

This study was financially supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government Ministry of Science and ICT (MSIP) [No. NRF-2020R1A4A4078916].

REFERENCES

[1] A. Waly, W. T.-A. in construction, and undefined 2003, “A virtual construction environment for preconstruction planning,” *Elsevier*.

[2] K.-J. Yi and D. Langford, “Scheduling-Based Risk Estimation and Safety Planning for Construction Projects,” *J. Constr. Eng. Manag.*, vol. 132, no. 6, pp. 626–635, Jun. 2006.

[3] S. J. Yoon, H. K. Lin, G. Chen, S. Yi, J. Choi, and Z. Rui, “Effect of Occupational Health and Safety Management System on Work-Related Accident Rate and Differences of Occupational Health and Safety Management System Awareness between Managers in South Korea’s Construction Industry,” *Saf. Health Work*, vol. 4, no. 4, pp. 201–209, Dec. 2013.

[4] X. Zheng, J. Yao, and X. Xu, “Violation Monitoring System for Power Construction Site,” *IOP Conf. Ser. Earth Environ. Sci.*, vol. 234, no. 1, p. 012062, Mar. 2019.

[5] A. K. Ali, O. J. Lee, and C. Park, “Near Real-Time Monitoring of Construction Progress: Integration of Extended Reality and Kinect V2,” Oct. 2020, doi: 10.22260/isarc2020/0005.

[6] S. Han and S. Lee, “A vision-based motion capture and recognition framework for behavior-based safety management,” *Autom. Constr.*, vol. 35, pp. 131–141, Nov. 2013.

[7] M. R. Teague, “IMAGE ANALYSIS VIA THE GENERAL THEORY OF MOMENTS.,” *J. Opt. Soc. Am.*, vol. 70, no. 8, pp. 920–930, 1980.

[8] H. Zhang, H. Z. Shu, P. Haigron, B. S. Li, and L. M. Luo, “Construction of a complete set of orthogonal Fourier-Mellin moment invariants for pattern recognition applications,” *Image Vis. Comput.*, vol. 28, no. 1, pp. 38–44, 2010.

[9] A. Bux, P. Angelov, Z. H.-A. in C. Intelligence, and undefined 2017, “Vision based human activity recognition: a review,” *Springer*.

[10] G. Mountrakis, J. Im, and C. Ogole, “ISPRS Journal of Photogrammetry and Remote Sensing Support vector machines in remote sensing: A review,” *ISPRS J. Photogramm. Remote Sens.*, 2010.

[11] Z. Miljković, B. Babic, N. Nesic, and Z. Miljkovic, “A review of automated feature recognition with rule-based pattern recognition Article in Computers in Industry · Chaotic metaheuristic algorithms for learning and reproduction of robot motion trajectories View project Nenad Nesic GE Power (Switzerland) 2 PUBLICATIONS 197 CITATIONS SEE PROFILE A review of automated feature recognition with rule-based pattern recognition,” 2008.

[12] J. Aggarwal, L. X.-P. R. Letters, and undefined 2014, “Human activity recognition from 3d data: A review,” *Elsevier*.

[13] S. Zhang, J. Teizer, J.-K. Lee, C. M. Eastman, and M. Venugopal, “Building Information Modeling (BIM) and Safety: Automatic Safety Checking of Construction Models and Schedules,” *Autom. Constr.*, vol. 29, pp. 183–195, Jan. 2013.

[14] K. Kim and Y. K. Cho, “Building Information Modeling for Temporary Structure Planning and Safety Analysis Construction Automation View project Construction Automation View project,” 2016.

[15] R. Haeusler and R. Klette, “Analysis of KITTI Data for Stereo Analysis with Stereo Confidence Measures,” Springer, Berlin, Heidelberg, 2012, pp. 158–167.

[16] S. Chi, S. Han, and D. Y. Kim, “Relationship between Unsafe Working Conditions and Workers’ Behavior and Impact of Working Conditions on Injury Severity in U.S. Construction Industry,” *J. Constr. Eng. Manag.*, vol. 139, no. 7, pp. 826–838, Jul. 2013.

[17] A. Montaser and O. Moselhi, “RFID indoor location identification for construction projects,” *Autom. Constr.*, vol. 39, pp. 167–179, Apr. 2014.

[18] Q. Fang *et al.*, “Detecting non-hardhat-use by a deep learning method from far-field surveillance videos,” *Autom. Constr.*, vol. 85, pp. 1–9, Jan. 2018.

[19] J. I. Messner, D. R. Riley, and M. J. Horman, “An Interactive Visualization Environment for Construction Engineering Education,” in *Construction Research Congress 2005*, 2005, pp. 1–10.

[20] G. H. Otto, J. I. Messner, and L. Kalisperis, “Expanding the Boundaries of Virtual Reality for Building Design and Construction,” in *Computing in Civil Engineering (2005)*, 2005, pp. 1–10.

[21] A. K. Ali, O. J. Lee, and H. Song, “Robot-based facade spatial assembly optimization,” *J. Build. Eng.*, vol. 33, p. 101556, Jan. 2021, doi: 10.1016/j.jobe.2020.101556.

[22] M. N. Vinodkumar and M. Bhasi, “Safety climate factors and its relationship with accidents and personal attributes in the chemical industry,” *Saf. Sci.*, vol. 47, no. 5, pp. 659–667, May 2009.

[23] H. Guo, Y. Yu, and M. Skitmore, “Visualization technology-based construction safety management: A review,” *Autom. Constr.*, vol. 73, pp. 135–144, Jan. 2017.

[24] V. Balali and M. Golparvar-Fard, “Segmentation and recognition of roadway assets from car-mounted camera video streams using a scalable non-parametric image parsing method,” *Autom. Constr.*, vol. 49, pp. 27–39, Jan. 2015.

[25] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement.”

[26] M. Menze, C. Heipke, and A. Geiger, “JOINT 3D ESTIMATION OF VEHICLES AND SCENE FLOW.”

[27] J.-R. Chang and Y.-S. Chen, “Pyramid Stereo Matching Network.”

[28] J. Seo, S. Han, S. Lee, H. K.-A. E. Informatics, and undefined 2015, “Computer vision techniques for construction safety and health monitoring,” *Elsevier*.

[29] M. Menze, C. Heipke, and A. Geiger, “Object Scene Flow.”