Enhancing Tower Crane Safety: A Computer Vision and Deep Learning Approach †

Parham Pazari [1], Nasim Didehvar [1] and Amin Alvanchi [1,*]

[1] Department of Civil Engineering, Sharif University of Technology, Tehran, Iran

* Correspondence: Alvanchi@sharif.edu

† Presented at the 1st International Online Conference on Buildings (IOCBD), Online, 24–26 October 2023

Abstract: The utilization of tower cranes at construction sites entails inherent risks, notably the potential for loads to fall. This study proposes a novel method for identifying the tower crane load fall zone and determining workers' locations relative to this zone. Previous studies have failed to accurately identify the load fall zone, mainly due to the difficulties in detecting various types of crane loads. This study presents a method that uses computer vision algorithms to detect crane loads based on their movement patterns and elevation, while also employing the YOLOv7 deep learning algorithm to detect workers and using stereo camera depth data to measure their positions in the 3D world coordinate system. The proposed method outperforms prior approaches in terms of analysis speed and accuracy, achieving a speed of 8 frames per second and 94% precision and 96.5% recall in determining workers' relative zone.

Keywords: Computer Vision, Deep Learning, Tower Crane, Construction Safety

1. Introduction

Construction is a high-risk activity that takes place in complex environments. Consequently, the construction industry experiences a high fatality rate. Many of these injuries and fatalities can be attributed to the use of cranes, which are extensively employed on construction sites [1]. In the United States, between 2011 and 2017, 297 fatal crane-related accidents were reported [2]. Among these incidents, 154 cases resulted from contact accidents, with 79 of them specifically involving the fall of crane loads [2]. Hence, the occurrence of crane load fall, accounts for more than 25% of all crane-related fatalities. Restricting access to hazardous crane areas is a fundamental measure that can be taken to prevent or mitigate the impact of loads falling [3].

The Occupational Safety and Health Administration (OSHA) defines the fall zone as the area including but not limited to the area directly beneath the load where there is a foreseeable possibility of suspended materials falling [4]. The OSHA 1926.1425 standard explicitly prohibits individuals from being present beneath a suspended load [4]. Similarly, the BS-EN-ISO-13857 standard recommends maintaining a safety distance of 1.5 meters from high-risk areas, including the crane load fall zone [5]. Automated safety systems, particularly those based on computer vision, have demonstrated promising potential for enhancing safety [6]. This study explores a computer vision and deep learning approach to accurately determine the relative location of individuals in relation to the tower crane load fall zone, which is validated through a carefully designed and executed experiment.

In the following sections of this paper, a review of the literature on crane safety improvement is presented in section 2. Section 3 provides an explanation of the methodology employed in this study. The experiment and its findings are presented in section 4. Section 5 is dedicated to the discussion of the results, and section 6 encompasses the conclusions drawn from the research.

2. Literature Review

Previous studies have focused on enhancing crane safety through the implementation of various technologies, including sensors, scanners, and computer vision [7]. However, in recent years, there has been a growing trend toward the use of computer vision technology for crane safety monitoring [6]. This literature review explores the research conducted in this area.

One of the objectives of previous studies has been to alleviate crane collisions with site entities. Zhang and Ge [8] proposed a deep learning algorithm, FairMOT, to predict the trajectories of individuals and crane loads in a 2D image space. Yang et al. [9] used the Mask R-CNN to detect individuals and crane hook. They calculate the distances between these entities to prevent collision. Chen et al. [10] examined the use of terrestrial laser scanning to produce a 3D point cloud and prevent crane collisions with modeled objects.

In addition to addressing crane collision prevention, computer vision technology has also been used to assist operators. Li et al. [11] developed an automatic system that uses robots and computer vision to attach loads to the crane hook. Wang et al. [12] trained a deep learning algorithm to interpret hand signals for crane steering. Other research [13, 14] has aimed to determine the precise location of the load during blind lifts, where the exact position of the load is unknown due to swaying. These studies used color-based identification techniques to detect the load. However, in real conditions, loads often have colors similar to the background, making color-based identification challenging. Yoshida et al. [15] demonstrated the feasibility of using a stereo vision system to detect the load location.

The identification of crane loads and their related hazardous areas has also been a topic of interest in previous research. Zhou et al. [16] used the Faster R-CNN algorithm to identify all objects that were in the shape of a cube as possible crane loads. Their research showed that the deep learning algorithms alone, was not able to distinguish between cuboid loads and other similar objects in the construction site. Chian et al. [17] proposed a method for estimating the tower crane load fall zone. Their method uses the homography matrix to transfer a grid of points from the project plan to the camera images. However, this method is only accurate if the construction site is a flat plane. In reality, the site is not a plane, which can cause parallax problem and errors in the identification of the fall zone. In order to solve the problems of previous research, the current research proposes a new method for monitoring the crane load fall zone and determining the presence of people under the load.

3. Methodology

The proposed algorithm consists of four main components, which are presented in the subsequent subsections: (1) depth extraction: A stereo camera system is used to extract depth information from the scene. (2) load detection: The crane load is recognized based on its movement patterns and elevation. (3) worker detection: The YOLOv7 algorithm is used to detect people in the scene. (4) location comparison: The location of people is compared to the location of the load fall zone in the scene. If a person is in the load fall zone, an alarm can be triggered to warn the person of the danger.

3.1. Depth extraction

Depth information is a necessary input for the load detection algorithm. It is also necessary to determine the location of people and the load fall zone in the 3D world coordinate system. Depth calculation using a stereo vision system is a well-established method for fast and accurate depth estimation [15]. Stereo vision is a technique that uses the difference between two or more stereo images to recover the three-dimensional structure of a scene [15]. The stereo vision system used in this study is a stereo normal case that consists of two cameras and provides disparity. Disparity is the difference in the image coordinates of the corresponding points in two stereo images. The relationship between disparity and depth is expressed in eq. (1) [18]. The disparity map was calculated using the StereoBM class of the OpenCV library [19].

$$\text{Disparity } = \text{x} - \text{x}' = \frac{\text{Bf}}{\text{z}}, \tag{1}$$

In eq. (1), x is the vertical coordinate of the point in the left image. x' is the vertical coordinate of the point in the right image. f is the focal length of the camera and B is the baseline distance between the two cameras and z is the depth of the point in three-dimensional space.

3.2. Load detection

Crane load detection is a more challenging task than regular object detection. The load is defined by the act of carrying it by the crane, rather than its visual appearance. Relying solely on the visual features of the load leads to limited results and makes it difficult to distinguish the load from other similar entities at the construction site [16]. In this study, a novel method for crane load detection was developed based on the movement pattern and height of the load. These two features define the load. As every load carried by a crane needs to be lifted from the surface below it and moved.

Detecting moving objects in a video is a common problem in computer vision [6]. However, detecting objects that move differently from the background is a different problem. When it is necessary to detect an object static with respect to the camera while the camera is moving, another method is needed. Optical flow is a two-dimensional vector that shows the movement of a specific point between two consecutive frames [20] and is a suitable criterion for identifying objects with different movements. The prerequisite for an object to be identified as a possible load is that it moves differently from the background. The optical flow in this study was calculated using the Lucas-Kanade algorithm [20] and then densely calculated using the interpolation technique for all pixels. By specifying pixels with different optical flow, objects with different motions can be identified.

The height difference between the load and the underlying surface is a crucial factor to consider. By comparing the depth of candidate objects and the surrounding surface within a radius of 1.5 times, it can be determined whether the object is suspended. A difference of 2 meters is considered the necessary distance for an object to be classified as suspended. When an object is suspended and exhibits different movement characteristics, it is considered as a load.

3.3. People detection

Identifying individuals from a high height with an almost vertical viewing angle is a challenging issue because classical algorithms depend on appearance features that are mostly absent [8]. In addition, the size of the people in the photos is very small. To overcome these challenges, this study employed the YOLOv7 algorithm [21], which was pre-trained on the COCO image database [22], to detect people with high accuracy.

Transfer learning of the YOLOv7 algorithm was performed using 120 images. Due to the small size of the database, the backbone layers of the pre-trained model, which are the first 50 layers, were frozen to avoid overfitting on a small amount of data.

3.4. Comparison of locations

The last step involves comparing the location of the individuals and the load fall zone. This study represents the first instance of such a comparison conducted in the 3D world coordinate system. Using eq. (2) and depth measurements, the 3D coordinate and true size of the objects can be determined, allowing for comparison in the world coordinate system.

$$\frac{x}{X} = \frac{y}{Y} = \frac{r}{R} = c \cdot \frac{f}{Z}, \tag{2}$$

Equation (2) employs the focal length (f), depth (z), and a constant for unit matching (c) to convert point image coordinates (x, y) to 3D world coordinates (X, Y), and the image size of objects (r) to its actual size (R).

Upon conducting the comparison of worker locations with the load fall zone, the worker locations are classified into three distinct zones: the red zone, which is directly beneath the load and is deemed off-limits to all personnel, the yellow zone, which is situated 1.5 meters away from the red zone and is only accessible to individuals involved in the operation, and the green/safe zone, which is situated outside these two zones.

4. Experiment

A field experiment was conducted to validate the proposed method, encompassing all loads handling scenarios at a height of 13 meters. To ensure that the test conditions closely resembled the actual working environment of the crane, a crane model was designed and fabricated. The length of the model boom was set at 3 meters to provide an adequate distance for the cameras from the load. Also, a stereo system comprising two calibrated smart phone cameras spaced at 18 cm was developed to capture videos.

The person's zone is the key variable and the ultimate outcome of this study. Precision and recall rate were used to assess the model's performance in determining the person's zone. The ground truth zone is documented by communicating with the individual under load and during the test execution. The confusion matrix related to determining the person's zone is presented in Table 1, which also reports the precision and recall rates for each zone. Figure 1 provides an example of the algorithm analysis results, revealing the precise detection of the load and the person. The person's zone is classified into three categories: red, yellow, and green, based on a comparison of their location with the fall zone's center point and the load's dimensions.

Table 1. Experiment confusion matrix and precision and recall rates for each zone.

Confusion matrix					Accuracy	
GT\Predicted	Red	Yellow	Green	Not	Precision (%)	Recall (%)
Red	175	1	0	11	97	94
Yellow	5	99	0	0	92	95
Green	0	8	234	0	100	97
Not	0	0	0	76	87	100
Average					94	96.5

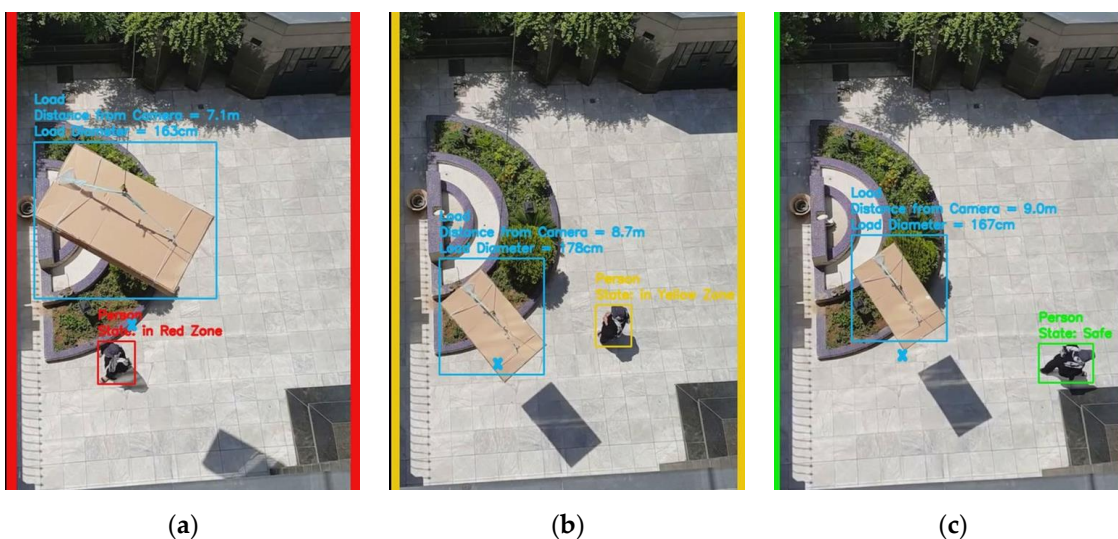(**a**)   (**b**)   (**c**)

Figure 1. Examples of the results of the experiment. (**a**) Person within fall zone; (**b**) Person is situated in the yellow zone and at a distance of one meter from the red zone; (**c**) Person located in a safe zone.

5. Discussion

The proposed model demonstrated high accuracy, as indicated by its 94% precision and 96.5% recall. The experiments were analyzed using a computer equipped with an

Intel Core i7-9750H CPU, 8 GB RAM, and NVIDIA GeForce GTX 1650 GPU. The algorithm runs at a speed of 8 frames per second. This execution speed determines the worker's zone almost in real time, making it suitable for monitoring the load fall zone and issuing a warning of worker presence in the red zone if needed.

The proposed method surpasses previous approaches in multiple key aspects. First, it is highly cost-effective because of its minimum physical equipment requirement. Second, it eliminates the reliance on unrealistic assumptions, such as assuming that the entire site surface is flat or that the camera is permanently fixed. This greatly enhances its practical use in construction sites. Third, the algorithm operates at a high speed of 8 frames per second, with an impressive 94% precision and 96.5% recall in detecting the worker's zone. This highlights the method's exceptional performance and reliability, surpassing previous models that identified only a limited set of load types [17] and a maximum speed of 1 frame per second [8]. Finally, the method's ability to provide continuous service without human intervention sets it apart from traditional methods that rely on human power for safety monitoring, reducing the likelihood of errors and improving overall efficiency.

Computer vision-based solutions have one limitation: the inability to function effectively in situations where the subject is occluded. This issue can be avoided to a large extent by determining the appropriate location for the cameras on the crane boom. However, in blind operations, occlusion may occur when the obstacle is very close to the load and its dimensions are significant compared with the height of the crane.

6. Conclusion

The present study introduces a novel approach to monitor the presence of individuals within the crane load fall zone, which is a critical issue in the construction industry due to legal emphasis and hazardous conditions. Crane safety has been the subject of extensive research. However, the practical application of past methods is restricted due to the cost and assumptions associated with them. This study proposes a method that uses stereo vision to extract image depth information, computer vision algorithms to detect the load, and the YOLOv7 deep learning algorithm to identify individuals. By comparing the location of individuals and the load fall zone in the world coordinate system, the zone of individuals is classified into three categories (red, yellow, and green) based on predefined rules. The key accomplishments of this research are as follows: (1) propose a novel approach for the detection of the crane load fall zone and the determination of workers' positions relative to it. (2) develop a new algorithm that enables load detection irrespective of its type, shape, color, and size. (3) fine-tune the YOLOv7 deep learning algorithm to accurately identify workers from the altitude of the crane. (4) construct a laboratory model of a crane to facilitate the testing of the proposed method.

In conclusion, the method presented in this study achieves 94% precision, 96.5% recall and an analysis speed of 8 frames per second, making it an accurate and fast solution that can be used in real-world conditions. This system can provide valuable information about the safe behavior of workers to safety managers.

References

1. Neitzel, R. L.; Seixas, N. S.; Ren, K. K., A review of crane safety in the construction industry. *Applied occupational and environmental hygiene* **2001,** *16* (12), 1106-1117.
2. Bureau of Labor Statistics, U. S. D. o. L. Fatal Occupational Injuries Involving Cranes. https://www.bls.gov/iif/oshwc/cfoi/cranes-2017.htm (accessed August 20).
3. Aneziris, O.; Papazoglou, I. A.; Mud, M.; Damen, M.; Kuiper, J.; Baksteen, H.; Ale, B.; Bellamy, L. J.; Hale, A. R.; Bloemhoff, A., Towards risk assessment for crane activities. *Safety science* **2008,** *46* (6), 872-884.
4. OSHA, O. S. a. H. A.-. Part 1926 - Safety and Health Regulations for Construction, Subpart CC - Cranes and Derricks in Construction. In *Title 29 - Labor, Subtitle B - Regulations Relating to Labor*, 2019; Vol. 1926.1425.
5. Standardization, E. C. f., Safety of machinery - Safety distances to prevent hazard zones being reached by upper and lower limbs. 2008.
6. Xu, S.; Wang, J.; Shou, W.; Ngo, T.; Sadick, A.-M.; Wang, X., Computer vision techniques in construction: a critical review. *Archives of Computational Methods in Engineering* **2021,** *28* (5), 3383-3397.
7. Sadeghi, S.; Soltanmohammadlou, N.; Rahnamayiezekavat, P., A systematic review of scholarly works addressing crane safety requirements. *Safety Science* **2021,** *133*, 105002.
8. Zhang, M.; Ge, S., Vision and trajectory–Based dynamic collision prewarning mechanism for tower cranes. *Journal of Construction Engineering and Management* **2022,** *148* (7), 04022057.
9. Yang, Z.; Yuan, Y.; Zhang, M.; Zhao, X.; Zhang, Y.; Tian, B., Safety distance identification for crane drivers based on mask R-CNN. *Sensors* **2019,** *19* (12), 2789.
10. Chen, J.; Fang, Y.; Cho, Y. K., Real-time 3D crane workspace update using a hybrid visualization approach. *Journal of Computing in Civil Engineering* **2017,** *31* (5), 04017049.
11. Li, H.; Luo, X.; Skitmore, M., Intelligent hoisting with car-like mobile robots. *Journal of Construction Engineering and Management* **2020,** *146* (12), 04020136.
12. Wang, X.; Zhu, Z., Vision-based hand signal recognition in construction: A feasibility study. *Automation in Construction* **2021,** *125*, 103625.
13. Fang, Y.; Chen, J.; Cho, Y. K.; Kim, K.; Zhang, S.; Perez, E., Vision-based load sway monitoring to improve crane safety in blind lifts. *Journal of Structural Integrity and Maintenance* **2018,** *3* (4), 233-242.
14. Price, L. C.; Chen, J.; Park, J.; Cho, Y. K., Multisensor-driven real-time crane monitoring system for blind lift operations: Lessons learned from a case study. *Automation in Construction* **2021,** *124*, 103552.
15. Yoshida, Y., Gaze-controlled stereo vision to measure position and track a moving object: Machine vision for crane control. *Sensing Technology: Current Status and Future Trends II* **2014**, 75-93.
16. Zhou, Y.; Guo, H.; Ma, L.; Zhang, Z.; Skitmore, M., Image-based onsite object recognition for automatic crane lifting tasks. *Automation in construction* **2021,** *123*, 103527.
17. Chian, E. Y. T.; Goh, Y. M.; Tian, J.; Guo, B. H., Dynamic identification of crane load fall zone: A computer vision approach. *Safety Science* **2022,** *156*, 105904.
18. OpenCV Depth Map from Stereo Images. https://docs.opencv.org/3.4/dd/d53/tutorial_py_depthmap.html (accessed 3, september).
19. OpenCV StereoBinaryBM Class Reference. https://docs.opencv.org/3.4/d7/d8e/classcv_1_1stereo_1_1StereoBinaryBM.html (accessed September, 2).
20. Lucas, B. D.; Kanade, T. In *An iterative image registration technique with an application to stereo vision*, IJCAI'81: 7th international joint conference on Artificial intelligence, 1981; pp 674-679.
21. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y. M. In *YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors*, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023; pp 7464-7475.
22. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C. L. In *Microsoft coco: Common objects in context*, Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13, Springer: 2014; pp 740-755.

Disclaimer/