Proceedings of the 2025 Winter Simulation Conference
E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.

3D VISION BASED ANTI-COLLISION SYSTEM FOR AUTOMATIC LOAD MOVEMENTS WITH TOWER CRANES - A SIMULATION ORIENTED DEVELOPMENT PROCESS

Alexander Schock-Schmidtke[1], Gonzalo Bernabé Caparrós[1], and Johannes Fottner[1]

[1]Chair of Materials Handling, Material Flow and Logistics,
School of Engineering and Design, Technical University of Munich, Munich, GERMANY

ABSTRACT

This paper presents a simulation-driven development approach for a camera-based anti-collision system designed for automated tower cranes. Stereo camera systems mounted directly on the crane's hook generate real-time 3D point clouds to detect people in the immediate danger zone of suspended loads. A virtual construction site was implemented in a game engine to simulate dynamic scenarios and varying weather conditions. The system utilizes a neural network for pedestrian detection and computes the minimum distance between load and detected persons. A closed-loop architecture enables real-time data exchange between simulation and processing components and allows easy transition to real-world cranes. The system was evaluated under different visibility conditions, showing high detection accuracy in clear weather and degraded performance in fog and rain due to the limitations of stereo vision. The results demonstrate the feasibility of using synthetic environments and point cloud-based perception to develop safety-critical assistance systems in construction automation.

1 INTRODUCTION

The construction industry remains one of the least automated industries, characterized by a high proportion of manual labor and limited integration of digital technologies. In contrast to sectors such as manufacturing or logistics, productivity in the construction industry has only increased by 1 % over the last two decades, despite increasing pressure due to the shortage of skilled labor, rising material costs, and shorter project deadlines (McKinsey 2020). The industry can counteract this obstacle primarily through digitalizing and automating construction machinery. On the one hand, this includes assistance systems for construction equipment that support workers in fulfilling their tasks, thereby increasing efficiency. On the other hand, fully autonomous construction machinery is being used to counteract the shortage of skilled labor and to meet ever-shorter process times. Automation in the construction industry is primarily focused on certain types of machines that are particularly suitable as autonomous systems due to their tasks and areas of application. As described in Schock-Schmidtke et al. (2024) and Nguyen and Ha (2023), this primarily includes machine types for earthmoving and transport, such as excavators, bulldozers, wheel loaders, or dump trucks. They perform repetitive tasks according to a cyclical principally pattern (first loading, then navigating to a defined position, and then unloading), which is why automation allows for a high increase in efficiency and rapid value creation. In civil engineering, the tower crane, in particular, significantly influences a construction site's productivity. It depends on the quality of crane operation by the user and the operation planning by the work scheduling department. Low efficiency and productivity can be attributed to the low utilization of the tower crane on the construction site, which is 50 % on average (Krause and Ulke 2016). The optimization of process time in civil engineering can therefore be achieved through a higher degree of automation of a tower crane. The conceptual implementation of a highly automated crane is described in Schock-Schmidtke et al. (2024). The paper presents an overall concept, including the system architecture for a highly automated tower crane that is able to move loads autonomously - with humans only as a fallback instance in the background. Such a system includes sensors that record the construction

site environment and software that generates a digital 3D construction site model. Suitable movement trajectories for load transport can be calculated based on this. However, continuous monitoring of the environment is required during the load movement in order to recognize and prevent possible collisions with people or other objects on the construction site at an early stage.

The development and validation of safety-critical assistance systems - such as an anti-collision system for autonomous cranes - requires the handling of potentially dangerous scenarios that can only be tested to a limited extent or not at all in a real environment. This is where game engines such as Unity or Unreal offer decisive advantages (Wolter 2021). Firstly, critical scenarios (e.g., collision with a moving pedestrian) can be easily reproduced in the simulation environment (Young et al. 2020). Secondly, the required sensor data, such as camera images or point clouds from LiDAR or radar, are often not available as real data but can be easily generated synthetically by game engines. Furthermore, many assistance systems rely on AI-based models that require a comprehensive annotated data set. The recording of data sets, including labelling in real environments, is time-consuming and cost-intensive (Endo et al. 2024). Game engines already offer automatic labeling and segmentation for this purpose. In summary, efficiency on the construction site can be increased through the use of automated tower cranes, among other methods. For this to be possible, the machines must be equipped with an intelligent anti-collision system.

This paper presents a camera-based anti-collision system that detects people in the immediate working area during the automated movement of a load and stops or slows down the crane movement in the event of an imminent collision. People are detected on the basis of point clouds created using stereo cameras in a virtual construction site environment. The simulation of the construction site and the sensor data is carried out in Unity. The rest of the paper is organized as follows: Section 2 reviews prior related work. In chapter 3, the problem is defined, and the research gap is derived from it. A detailed description of the simulation model and the developed anti-collision model is provided in Section 4 and 5. The paper lists the simulation results in Section 6 followed by a concluding discussion and outlook on future work steps in Section 7.

2 RELATED WORK

The development of safety-critical assistance systems for automated construction machinery requires an interdisciplinary approach that combines methods from robotics, computer vision, and simulation. In recent years, both simulation technologies and sensor-based anti-collision systems have seen significant advancements. This chapter reviews the current state of the art in two key areas relevant to this work: the use of simulation environments for system development, synthetic data generation, and validation; and existing anti-collision systems for tower cranes. The aim is to highlight current research gaps that motivate the proposed approach.

2.1 Simulation and Virtual Environments

The use of simulation to model the kinematics and dynamics of machines and vehicles is a widespread approach in the product development process. However, additional information, such as the reconstruction of the environment, the consideration of environmental influences, or the reception of raw sensor data, is important for the realization of autonomously acting systems. Various simulation tools and approaches are already being used in the automotive industry for this (Rong et al. 2020). They offer significantly more advantages compared to real test series. Real tests are cost-intensive, cannot be repeated at will, are difficult to control and are often associated with risks. In addition, they only represent a limited number of possible test scenarios. Furthermore, conventionally recorded test data from multi-sensor systems does not enable closed-loop testing, i.e., there is no feedback between sensor and system reaction. This is where 3D simulations offer major advantages: they are scaleable, generic, more cost-effective, and allow any scenario to be run through in a targeted manner. They also automatically provide the required ground truth information (i.e. the 'true' state of the environment) for AI training data or even completely automatically

labeled data sets. In the virtual environment, a digital representation of the system to be developed is confronted with dynamic scenes (e.g. changing weather conditions or changing work areas). The virtual sensor system generates camera data, LiDAR points or radar data, taking into account the physical properties of bodies, as well as their material/surface properties and the lighting conditions in the simulated world. The aim is to make this environment as realistic as possible and, at the same time, simulate it in real time in order to integrate it into software-in-the-loop or hardware-in-the-loop test systems, for example. The simulation is particularly effective if it is closed-loop, i.e. the behavior of the system in turn influences the environment. This is not the case with open-loop tests, where only individual system states can be tested like mentioned in Wolter (2021) and Kiran et al. (2024).

The use of simulation environments to support the development of construction machinery is not as advanced as it is in other industries. Current research results on the use of simulation in construction robotics and automation are shown in (Xu et al. 2024). Accordingly, game engines are primarily used to validate algorithms and models in a virtual construction site and for visualization. In their work, Pereira Da Silva et al. (2022) demonstrate the use of a simulation in Unity to recognize potential construction problems at an early stage and to evaluate construction processes in terms of duration and costs. The game engine is primarily used here to realistically visualize the construction process and to test alternative execution scenarios under controlled conditions. For process planning of the construction site with deep reinforcement learning, Zhu et al. (2023) also use a game engine. Realistic construction site scenarios are provided in the virtual environment in order to train and validate autonomous decision-making processes safely and efficiently. With a greater focus on implementing long-horizon tasks, Huang et al. (2023) use simulation models to test reinforcement learning-based construction robotic control agents and evaluate their performance. Consequently, when using game engines for simulation purposes, the focus is on validating algorithms and control systems. They enable efficient and low-risk validation of control strategies in realistic test environments and aim to visualize complex processes. Initial approaches to using simulation models to generate synthetic sensor data are showing promising results, particularly in research on construction robotics. For example, Aluckal et al. (2025) developed a simulation environment for autonomous construction machinery in the TERA project and Endo et al. (2024) in the open-source OPERA project. Furthermore, the simulation projects allow detailed modeling of the interaction between the machine and the deformable ground using physics engines and, thus, the synthetic determination of the machine kinematics. On the other hand, simulation also enables the integration of virtual sensors (cameras, IMUs, LiDAR) and the simultaneous simulation of several machines. This enables the efficient generation of large, annotated data sets for the evaluation of AI-based control strategies. In summary, it can be stated that game engines are not only used in current research for visualization but primarily as powerful tools for generating synthetic training data for AI-based systems in construction robotics.

2.2 Anti-Collision Systems for Tower Cranes

From a technical perspective, anti-collision systems for tower cranes are used today, particularly where several cranes operate in overlapping work areas or where structural obstacles and restricted visibility increase the risk of collisions. Regardless of the manufacturer like (CAD.42 2024), (AMCS technologies 2024) or (SMIE 2025), these systems follow a common basic principle, which largely coincides in the central functions. The safety systems comprise a combination of position and motion detection, real-time monitoring, and automated intervention logic. The typical system architecture comprises sensors at the machine's degrees of freedom (e.g., rotary encoders or inclination sensors), a central computing unit for motion analysis, and a human-machine interface (HMI) for visualization and input. The sensors continuously record the crane's current kinematics and derive the crane hook's position from this. From this, the position of the crane in a 3D space can be determined, which is visualized for the crane operator using the HMI. The user interface can also be used to statically implement exclusion zones and safety distances to defined objects. The anti-collision system first warns the operator in the event of an impending collision between cranes or when entering a restricted zone and, in an emergency, actively intervenes in the crane control

system and stops the load movement. Apart from anti-collision systems to protect against collisions between cranes, the use of safety systems that warn of and protect against collisions between loads on the crane and people or objects in the vicinity is not yet known.

Nevertheless, there are already some scientific approaches that deal with the topic of human protection/collision warning during load transport with cranes. As Ali et al. (2024) summarize, the detection of the construction site environment during load transport is carried out either via point clouds using LiDAR and radar or by means of camera images, including object detection. Both Yong et al. (2023) and Ku et al. (2024) present approaches for collision avoidance during crane operation that are based on sensor-based detection of the environment (point clouds) and subsequent processing of this data. The point clouds enable the exact localization of objects in relation to the load on the crane hook. Yong et al. combine a camera with a deep learning algorithm (Faster R-CNN) and ultrasonic sensors to recognize people and objects in the crane environment. Ku et al., on the other hand, use a crane system with LiDAR and radar sensors to detect obstacles. Both systems share the basic idea that multi-modal sensor data and AI-based environmental perception form an essential basis for safe, automated lifting processes. However, the sensors always look vertically at the load and only cover the near-field area directly around the load. Camera-based solutions are shown by Yang et al. (2019) and Pazari et al. (2023). The work by Yang et al. (2019) shows how the use of Mask R-CNN can be used to develop a camera-based system for the automated detection of people and danger zones under a tower crane. The distance between the crane hook, loads and workers is calculated from the pixel coordinates and the technical specifications of the camera sensor, which is located on the crane's trolley. This means that no further sensors are required on the crane to detect the surroundings. However, the quality of the measurements suffers if camera visibility is poor (e.g. due to bad weather conditions) or if the position of the camera changes due to vibrations of the crane boom. Pazari et al. present a system for automated detection of the danger zone under suspended loads of tower cranes. The work aims to identify construction workers in real time and categorize them in relation to this fall zone to avoid collisions and accidents. The developed anti-collision system is based on a generic visual analysis using deep learning and stereo-image processing. First, the depth is measured with a stereo camera, then a pre-trained network is used to detect people in the danger zone from the RGB images and then transferred to a 3D model. However, this approach has weaknesses, as the stereo camera is mounted on the trolley, meaning that the system can only provide an accurate resolution for a limited range. Furthermore, as tower cranes often have to lift the load to be transported over buildings, there is a risk that the camera system will not be able to capture the load and its surroundings.

In the field of anti-collision systems for tower cranes, it is clear that the current focus is primarily on avoiding collisions between cranes. Systems for detecting and avoiding hazards caused by the suspended load in combination with people in the work area are rarely used in industry. In science, concepts have already been presented that have potential, but do not represent a sufficient solution due to a field of vision that is too small or a lack of reliability.

3 PROBLEM DESCRIPTION

As can be seen from Section 2, the current focus of existing anti-collision systems is primarily on preventing collisions between several cranes or between cranes and static obstacles. However, the area under and around the load - the so-called drop zone - has hardly been taken into account to date, although this is precisely where uncontrolled movements of the load pose a high potential risk to people. Furthermore, such a system is essential for automated load transport with cranes, as the machine requires information about the working environment and the required safety distances. However, existing camera-based or sensor-fusion research approaches aim at recognizing people and objects near the suspended load are often limited by the viewing angle or sensor range. The research approaches presented rely on camera-based systems with object detection, e.g. based on YOLO architectures. These usually work with top-down perspectives, in which a camera on the trolley or boom looks downwards. However, this method has several limitations.

At great working heights, people only appear very small in the image. Often, only the helmet is visible as a colored circle, which means that reliable recognition is no longer possible.

To overcome these limitations, an alternative solution is proposed in this paper. Multiple stereo camera systems are mounted directly on the crane's hook. The camera systems are arranged offset to each other and have overlapping fields of view. This allows a large three-dimensional detection area around the load to be monitored, which is visualized in the form of point clouds. These point clouds then serve as the basis for recognizing people using deep learning. In contrast to purely 2D-based methods, point clouds provide a geometric representation of the objects in the room - so the shape of a person can be explicitly recognized and localized. The advantage of this approach is that the working range around the load can be better determined because the sensor is mounted on the hook. On the other hand, people can be better identified using envelope shapes (point clouds) than with object detection using 2D images. The described anti-collision system is developed with the help of a game engine, which depicts different construction site events and provides images from the stereo cameras.

4 MODEL SETUP

The development and validation of safety-critical assistance systems - such as an anti-collision system for autonomous cranes - requires handling potentially dangerous scenarios that can only be tested to a limited extent or not at all in a real environment. This is where the use of game engines to simulate scenarios and generate synthetic data offers decisive advantages. The following Figure 1 shows the system architecture of the simulation-based approach. This is a closed-loop simulation consisting of the Unity simulation

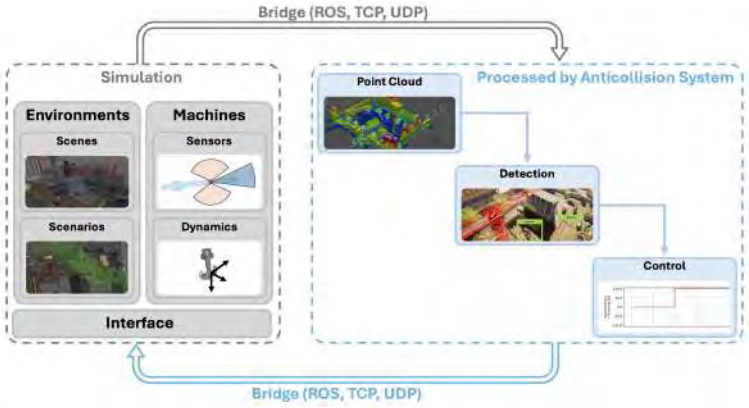


Figure 1: High-level architecture of the anti-collision system and the usage of closed-loop simulation with game engines.

environment, which represents a virtual construction site with cranes, and an external computing unit, which calculates and implements the anti-collision function. Data is exchanged between the simulation and the anti-collision system processing unit via a data bridge (ROS or TCP/UDP protocol). The system architecture was intentionally designed to allow an easy transition from a simulated to a real-world setup. By separating the simulation environment (Unity) from the processing unit and connecting both via ROS, the simulation can later be replaced by real hardware without changing the core communication or processing logic. This modular setup enables seamless deployment of the anti-collision system on an actual tower crane, using real sensor inputs in place of the simulated ones. As a result, the development environment supports rapid prototyping and hardware-in-the-loop testing with minimal adaptation effort.

4.1 Virtual Representation of the Construction Site

The simulation model represents a high-rise building with dynamic construction site activity. The High Definition Render Pipeline (HDRP) in Unity is used to realize the environment. It enables high-resolution, physically based modeling and representation of materials, light and environmental influences, allowing realistic simulation results and photo-realistic images to be created. The functionality of the simulation environment essentially comprises the two core areas, environments and machines (ref. Figure 1). The environment simulation includes the representation of a construction site (ref. Figure 2), which is based on the classic structure or layout according to (Schach and Otto 2011). In addition to the objects to be

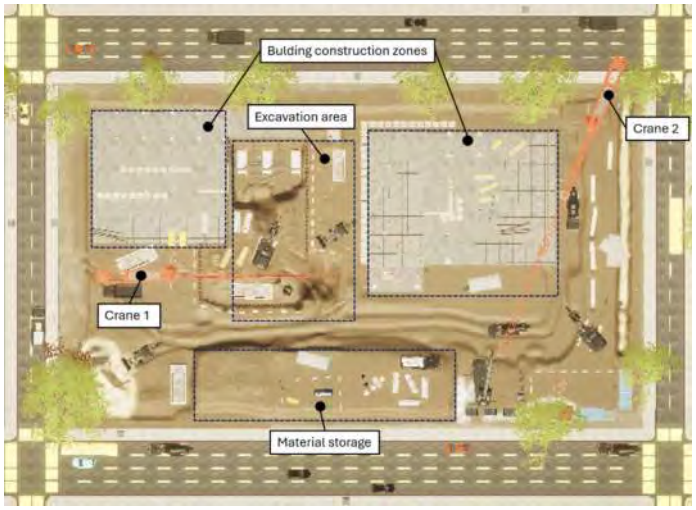


Figure 2: Rendering example of the construction site scene in unity.

built (concrete buildings), the construction site layout also includes storage areas for construction aids (e.g., formwork panels, scaffolding, etc.), a zone for storing construction materials such as sand and gravel, and an area for containers and recreation rooms. There is also a delivery zone and unloading areas for trucks. The construction site is separated from the surrounding urban environment by fences. The surface texture was created using the Unity Terrain Tool so that the simulated construction site has realistic ground characteristics such as ruts, uneven ground and material properties. The construction site model also includes a large number of construction machines, construction workers and components of a construction site infrastructure. As the simulation is a dynamic representation, machines and vehicles such as trucks or excavators can be defined as moving objects that move along a trajectory or carry out a work task (e.g. digging a trench). The construction site workers also move around the construction site, but not according to a fixed pattern, but rather according to a random principle so that no patterns are recognizable in the subsequent generation of synthetic data. For a realistic representation of the environment, the model takes into account both the lighting depending on the time of day and the current weather conditions (i. e. sunshine, cloudy skies, fog and rain) and consequently also all physical effects such as gravity. As the assistance system to be developed is designed for cranes, the tower crane in particular was implemented as an interactively controllable system. This means that each degree of freedom of the machine (rotation of the boom by means of the slewing gear, movement of the trolley along the boom, lifting and lowering of the load) can be individually controlled and regulated so that the kinematic behavior corresponds to that of a real machine.

4.2 Sensor Setup

The anti-collision system developed is based on a combination of several stereo-optical sensor systems that are mounted directly on the hook block of the tower crane. The aim of this configuration is to precisely

detect the danger zone around the suspended load and to enable robust object detection - especially of people. Figure 3 shows a schematic arrangement of the cameras and the corresponding coverage area.

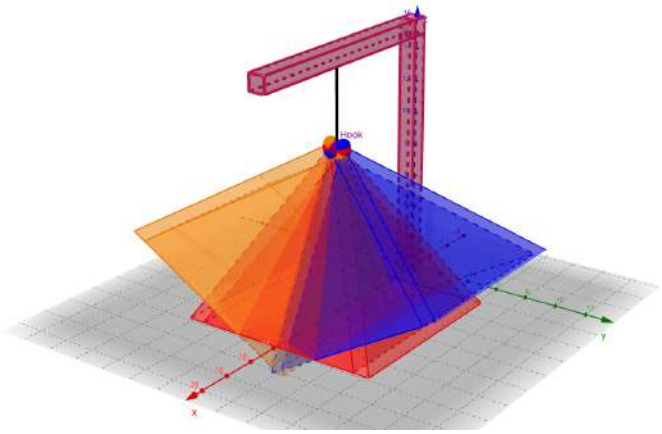


Figure 3: Schematic field of view of the stereo cameras on the crane hook.

Accordingly, a central stereo camera looking vertically downward captures the immediate danger zone directly around the load. Two further stereo camera systems are attached to the left and right of the hook and mounted at an angle of 45 degrees to the vertical to ensure an extended detection area around the load. This configuration creates a cylindrical field of view that can capture both the immediate close-up area and lateral sources of danger - such as approaching persons or vehicles. The technical characteristics of the cameras are shown in Table 1. The sensors have a resolution of 3840 x 2160 pixels (8.3 megapixels) and a vertical field of view (FOV) of 60 °. This results in a horizontal FOV of approx. 91 °, based on the aspect ratio of 16:9. For the assistance system, a high-depth resolution in the vicinity of the load is particularly relevant. A depth resolution of 50 mm at a distance of 10 m is defined as the target for the system. The diagram in Figure 4 shows the depth resolutions as a function of different baselines in the distance range up to 20 m. A baseline of 1 m best meets this requirement. According to (Hartley and Zisserman 2003), the minimum measurable depth at a baseline of 1 m and a maximum disparity d_{max} of 1920 pixels (half the image width) is 1.73 m. In addition to mechanical integration and optical alignment, the cameras are

Table 1: Technical specification of the stereo camera system.

resolution	3840 px x 2160 px (8.3 MP)
field of view	60 °
baseline	1 m
disparity accuracy	1 px
minimal depth	1.73 m
depth resolution at 10 m	53.5 mm

integrated into the system by software. Each stereo camera is provided with its own IP address and can send its raw data to central processing units for further processing via the TCP protocol. To illustrate the functioning under different environmental conditions, Figure 5 shows exemplary raw images of a camera in different weather scenarios. The examples include sunny conditions, fog and rain – typical situations that may occur during subsequent construction site operation and affect the performance of stereo depth detection.

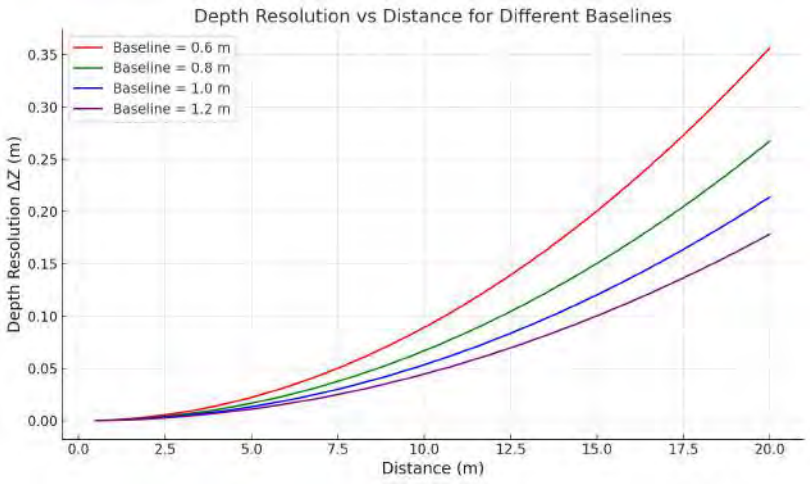


Figure 4: Comparison of the depth resolution for different baselines.

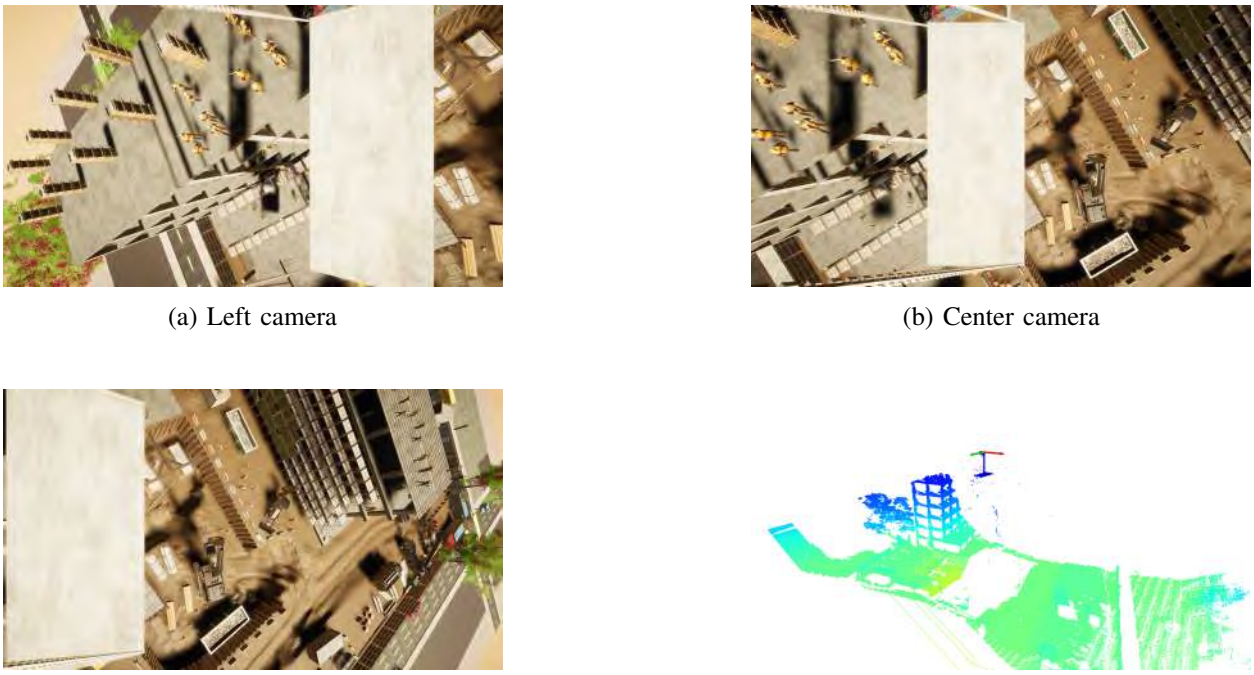
(a) Sunny weather (b) Foggy environment (c) Cloudy weather (d) Rainy weather

Figure 5: Examples of raw images from the stereo camera in various weather conditions.

5 POINT-CLOUD BASED ANTI-COLLISION SYSTEM

The anti-collision system developed is based on the analysis and interpretation of three-dimensional point clouds that are generated in real-time by several stereo cameras. The aim is to reliably detect people and objects in the danger zone around the suspended load and to initiate automated measures in the event of an impending collision. Before the camera data can be used, the stereo cameras must be calibrated to determine the exact geometric relationship between the two lenses. Chessboard patterns with 28 x 19 squares are used for this purpose, which allows the intrinsic and extrinsic parameters such as the translation vector, the rotation vector and the distortion coefficients of the sensors to be calculated. The reprojection error measures the difference between an observed point in the image (e.g., a corner of a chessboard pattern) and the point calculated by the camera model after reprojecting the 3D world point back into the 2D image. The goal is to achieve a reprojection error < 0.5 pixel because the smaller this value, the higher the accuracy of the subsequent point cloud. In summary, it can be said that the quality of the calibration has a significant influence on the subsequent measurement accuracy, which is why a large diversity of calibration images (the checkerboard is placed with a different orientation and at a different distance from both cameras to create a higher variety of conditions) is desirable. Image rectification is carried out after calibration. Two different perspective images from a stereo camera system are geometrically transformed so that corresponding image points lie on the same line (epipolar line). This process reduces the computational effort for the subsequent calculation of the depth map of each stereo image pair since the corresponding points can now be found more easily. The synchronized stereo image pairs are then used to generate disparity maps, from which depth maps are calculated using the camera parameters. This depth information is converted into a three-dimensional point cloud format using the camera parameters and the stereo semi-global block matching (SGBM) algorithm. Outliers are then determined using KD trees and the nearest neighbor method. This helps to suppress measurement errors and noise effects, making

the point cloud smaller and improving its quality. Finally, the individual point clouds of the three stereo camera systems are integrated into a global coordinate system by a coordinate transformation. Due to the overlap of the FOV of the individual stereo cameras, the density of the global point cloud increases so that a high-resolution reconstruction of the environment is possible. The result of an example recording consisting of an RGB image for each of the three cameras and the point cloud calculated from it is shown in Figure 6.



(a) Left camera

(b) Center camera

(c) Right camera

(d) Point cloud

Figure 6: Example of point cloud based on a stereo camera.

The PointPillars network was used to detect people in the point clouds generated by stereo cameras. PointPillars is an efficient 3D detector that was originally developed for LiDAR-based point clouds but can also be applied to other types of 3D data. The approach is characterized by particularly fast inference speed combined with high detection accuracy and was, among other things, specifically trained to detect pedestrians (Lang et al. 2019). The fused point cloud from the three stereo camera systems is used to determine the minimum distance between a detected person and the suspended load. After successful registration of the point clouds in a common coordinate system, the load is first identified within the point cloud. Since the camera systems are mounted on the hook block and primarily work in a top-down perspective, the load is usually located in the center of the captured field of view. A segment-based analysis – based on point density, height distribution and compactness – can be used to reliably delimit the point cloud of the load. At the same time, the PointPillars network detects people. The global coordinate system enters the detected people as 3D bounding boxes. The minimum Euclidean distance between the point clouds or their circumscribing bounding volumes is then calculated to determine the distance between the load and the person. For this purpose, a KD-tree-based nearest neighbor search is used to efficiently determine the smallest distance between the relevant point groups. The system remains passive if the calculated distance exceeds a defined safety limit – e.g., 2.0 m. However, if the distance falls below this threshold, the movement of the load is slowed down or stopped completely. The critical distance can be configured for each specific project and depends on factors such as the weight of the load, the speed of movement and the visibility. By transmitting the distance values to the simulation and the integrated

crane control, the simulation cycle is closed again (closed-loop simulation) and an overall evaluation of the system is made possible.

6 EXPERIMENTAL RESULTS

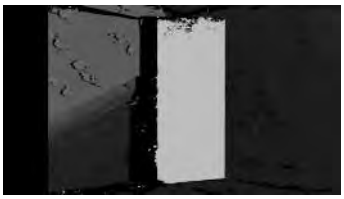
To validate the developed anti-collision system, the focus is particularly on the reliability of human detection under realistic environmental conditions. The distance between the load and the detected person is determined based on the 3D point cloud and, from a technical point of view, represents a geometrically trivial calculation. The real challenge lies in the reliable detection of persons, especially in adverse weather conditions, which directly impact the quality of the point cloud. To examine the quality of the 3D point clouds and the detection of people, 25 exemplary scenarios were carried out in the simulation environment with three different weather scenarios (sunny weather with good visibility, fog with reduced visibility and diffuse light conditions, and rain with poor visibility). In the construction site scenarios, a concrete floor slab is transported by a crane to different work areas. There are always several people in the work area and near the load.

(a) Sunny RGB image

(b) Foggy RGB image

(c) Rainy RGB image

(d) Disparity image in sunny weather



(e) Disparity image in foggy weather

(f) Disparity image in rainy weather

Figure 7: Visualization of RGB and disparity images under different weather conditions (top: RGB, bottom: disparity).

Exemplary shown in Figure 7 are the camera recordings of a stereo camera looking straight down and the corresponding disparity image for a construction site scenario. It can be seen that a clean, homogeneous disparity is present when recording under sunny conditions. In addition, the edges are hardly frayed and the object contours are clearly and sharply visible. In comparison, the image taken in fog is of significantly poorer quality, resulting in the edges in the disparity image being more frayed and the image showing many small artifacts. When it rains, isolated artifacts caused by light reflections appear. The central structure remains largely intact, but the object edges are softer and less defined. For the three different weather scenarios, the respective point clouds are generated from the disparities and then the person detection is carried out using PointPillars. The results are listed in Table 2. The confidence score is used to evaluate the person's identification and is averaged over the test series. These results make it clear that the quality of the depth images – and thus of the point clouds – depends to a large extent on external influences. Weather conditions have a strong impact on the visual sensors and lead to measurable deviations in the detection performance of the system.

Table 2: Effects of weather conditions on the detection of people.

weather condition	**mean confidence score**	**qualitative evaluation**
sunny	0.88	very high quality, clean point clouds, complete person shape
fog	0.53	partially fragmented point clouds, missing body parts, especially in the lower area (e.g., legs), lower point density
rain	0.71	drops create artifacts, point clouds are noisy, but the shape of people is usually still recognizable

7 DISCUSSION AND CONCLUSIONS

The simulation results confirm that the developed anti-collision system is capable of detecting people within the danger zone of a suspended load, particularly under favorable visibility conditions. In sunny weather, the generated 3D point clouds were complete and accurate, yielding a high average confidence score of 0.88. However, in challenging conditions such as fog or rain, detection performance decreased due to physical limitations inherent to stereo vision, such as reduced contrast and optical noise. These limitations led to fragmented point clouds, blurred object contours, and reduced confidence in the detection algorithm. Despite this, the camera positioning directly on the hook block, combined with overlapping fields of view, significantly improved the robustness of the detection range compared to traditional top-down or static camera systems. Furthermore, the system reliably calculated minimum distances in 3D space across all test conditions, highlighting the potential for safe real-time intervention in crane operations. A key strength of the simulation approach is its modular simulation architecture, which allows seamless replacement of the virtual environment with real hardware—enabling hardware-in-the-loop testing and easy deployment on actual cranes.

However, several limitations remain. Most notably, the system's sensitivity to environmental conditions must be addressed. While the current implementation demonstrates feasibility, it is not yet robust enough for reliable use in all weather scenarios. Furthermore, the evaluation so far has focused on simulated scenes; real-world experiments are needed to validate system performance under uncontrolled, dynamic site conditions. Future research should focus on developing a custom neural network architecture specifically tailored to stereo-based point cloud data, capable of handling noisy, sparse, and partial inputs. Additionally, extending the system to detect other hazard sources (e.g., moving machinery, structural obstacles) could further increase safety. A comparison with traditional overhead systems could further validate the benefits of the hook-mounted setup. The simulation-based workflow using Unity proved effective for safe, rapid prototyping and shows strong potential for developing AI-based safety systems in construction.

REFERENCES

Ali, A. H., T. Zayed, R. D. Wang, and M. Y. S. Kit. 2024. "Tower Crane Safety Technologies: A Synthesis of Academic Research and Industry Insights". *Automation in Construction* 163:105429.

Aluckal, C., R. V. Kumar Lal, S. Courtney, Y. Turkar, Y. Dighe, Y. Kim, *et al*. 2025. "TERA: A Simulation Environment for Terrain Excavation Robot Autonomy". In *2025 IEEE International Conference on Simulation, Modeling, and Programming for Autonomous Robots (SIMPAR)*, 1–6.

AMCS technologies 2024. "Antikollisionssysteme". https://amcs.fr/de/, accessed 29th March 2025.

CAD.42 2024. "Antikollisions-Turmdrehkrane / Mobilkrane". https://cad42.com/de/securite/Anti-Kollisions-Turmkrane-Mobilkr%C3%A4ne/, accessed 29th March 2025.

Endo, D., Y. Matsusaka, G. Yamauchi, and T. Hashimoto. 2024. "Research on an Open Source Physical Simulator for Autonomous Construction Machinery Development". In *41st International Symposium on Automation and Robotics in Construction*, Lille, France, 1303–1306.

Hartley, R., and A. Zisserman. 2003. *Multiple View Geometry in Computer Vision*. 2nd ed ed. Cambridge, UK ; New York: Cambridge University Press.

Huang, L., Z. Zhu, and Z. Zou. 2023. "To Imitate or Not to Imitate: Boosting Reinforcement Learning-Based Construction Robotic Control for Long-Horizon Tasks Using Virtual Demonstrations". *Automation in Construction* 146:104691.

Kiran, A., S. S. Salunkhe, B. Srinivas, M. Vanitha, M. Altaf Ahmed, and J. Venkata Naga Ramesh. 2024. "Accelerating Autonomous Vehicle Safety through Real-Time Immersive Virtual Reality Gaming Simulations". *Entertainment Computing* 50:100674.

Krause, T., and B. Ulke. (Eds.) 2016. *Zahlentafeln für den Baubetrieb*. Wiesbaden: Springer Fachmedien Wiesbaden.

Ku, T. K. X., B. Zuo, and W. T. Ang. 2024. "Robotic Tower Cranes with Hardware-in-the-Loop: Enhancing Construction Safety and Efficiency". *Automation in Construction* 168:105765.

Lang, A. H., S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom. 2019. "PointPillars: Fast Encoders for Object Detection From Point Clouds". In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 12689–12697.

McKinsey 2020. "The next Normal in Construction - How Disruption Is Reshaping the World's Largest Ecosystem". https://www.mckinsey.com/~/media/McKinsey/Industries/Capital%20Projects%20and%20Infrastructure/Our%20Insights/The%20next%20normal%20in%20construction/The-next-normal-in-construction.pdf, accessed 29th March 2025.

Nguyen, H. A., and Q. P. Ha. 2023. "Robotic Autonomous Systems for Earthmoving Equipment Operating in Volatile Conditions and Teaming Capacity: A Survey". *Robotica* 41(2):486–510.

Pazari, P., N. Didehvar, and A. Alvanchi. 2023. "Enhancing Tower Crane Safety: A Computer Vision and Deep Learning Approach". *Engineering Proceedings* 53(1).

Pereira Da Silva, N., S. Eloy, and R. Resende. 2022. "Robotic Construction Analysis: Simulation with Virtual Reality". *Heliyon* 8(10):e11039.

Rong, G., B. H. Shin, H. Tabatabaee, Q. Lu, S. Lemke, M. Možeiko, *et al*. 2020. "LGSVL Simulator: A High Fidelity Simulator for Autonomous Driving". In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 1–6.

Schach, R., and J. Otto. 2011. *Grundlagen der Baustelleneinrichtungsplanung*. Wiesbaden: Vieweg+Teubner.

Schock-Schmidtke, A., T. Bernhard, and J. Fottner. 2024. "Concept of a Highly Automated Tower Crane". In *2024 IEEE 20th International Conference on Automation Science and Engineering (CASE)*, Bari, Italy, 4062–4067.

SMIE 2025. "ProSITE". https://smie.com/prosite-2/, accessed 29th March 2025.

Wolter, S. 2021. "Simulation von Sensordaten mittels Game Engines". *ATZextra* 26(S3):24–27.

Xu, L., H. Liu, B. Xiao, X. Luo, DharmarajVeeramani, and Z. Zhu. 2024. "A Systematic Review and Evaluation of Synthetic Simulated Data Generation Strategies for Deep Learning Applications in Construction". *Advanced Engineering Informatics* 62:102699.

Yang, Z., Y. Yuan, M. Zhang, X. Zhao, Y. Zhang, and B. Tian. 2019. "Safety Distance Identification for Crane Drivers Based on Mask R-CNN". *Sensors* 19(12):2789.

Yong, Y. P., S. J. Lee, Y. H. Chang, K. H. Lee, S. W. Kwon, C. S. Cho *et al*. 2023. "Object Detection and Distance Measurement Algorithm for Collision Avoidance of Precast Concrete Installation during Crane Lifting Process". *Buildings* 13(10):2551.

Young, P., S. Kysar, and J. P. Bos. 2020. "Unreal as a Simulation Environment for Offroad Autonomy". In *Autonomous Systems: Sensors, Processing, and Security for Vehicles and Infrastructure 2020*, Online Only, United States.

Zhu, A., T. Dai, G. Xu, P. Pauwels, B. De Vries, and M. Fang. 2023. "Deep Reinforcement Learning for Real-Time Assembly Planning in Robot-Based Prefabricated Construction". *IEEE Transactions on Automation Science and Engineering* 20(3):1515–1526.

AUTHOR BIOGRAPHIES

ALEXANDER SCHOCK-SCHMIDTKE is a Research Associate and Ph.D. candidate at the Chair of Materials Handling, Material Flow and Logistics at the Technical University of Munich. He received his M.Sc. degree in Mechanical Engineering from TUM. His research focuses on the automation of construction machinery, with an emphasis on perception systems, navigation algorithms, and sensor fusion for safety-critical applications. His email address is alexander.schock@tum.de.

GONZALO BERNABÉ CAPARRÓS is a Master's student in the School of Engineering and Design at Technical University of Munich, Germany. His academic interests include simulation-based development methods, and robotics. His email address is g.bernabe@tum.de.

JOHANNES FOTTNER received a Dr.-Ing. degree in mechanical engineering from the Technical University of Munich (TUM), Munich, Germany in 2002. From 2002 to 2016, Johannes Fottner acted in different managing functions in the materials handling sector and gained expertise in the development, implementation and operation of automated logistics and production systems composed of different modules. Since 2016, he has been the head of the Chair of Materials Handling, Material Flow, Logistics at the TUM. His current research interests include investigating innovative technical solutions and system approaches for optimizing logistical processes. His email address is j.fottner@tum.de.