ICCEPM 2024 https://dx.doi.org/10.6106/ICCEPM.2024.0629
The 10th International Conference on Construction Engineering and Project Management
Jul. 29- Aug. 1, 2024, Sapporo

Enhancing On-Site Construction Machinery Handling through 3D Spatial Gesture-Based Trajectory Interpreter Modeling

Du LI [1]*, Ying WANG [2], Hung-Lin CHI [3]
[1] *Department of Building and Real Estate, Faculty of Construction and Environment, The Hong Kong Polytechnic University, Hong Kong*, E-mail address: 928renli@gmail.com
[2] *Department of Building and Real Estate, Faculty of Construction and Environment, The Hong Kong Polytechnic University, Hong Kong*, E-mail address: yingzc.wang@polyu.edu.hk
[3] *Department of Building and Real Estate, Faculty of Construction and Environment, The Hong Kong Polytechnic University, Hong Kong*, E-mail address: hung-lin.chi@polyu.edu.hk

Abstract:

In construction projects, the safety and productivity of machinery operations are of paramount importance. Contemporary research and industry endeavors predominantly concentrate on equipping machine operators with sensory information and establishing a comprehensive situation-aware operating environment, such as virtual reality-based machine manipulation training. However, significant limitations exist in direct information exchange and processing by on-site personnel. Notably, research on analyzing communication patterns in construction machinery operations remains scarce despite its critical role in preventing hazardous instructions/actions and enhancing machinery work efficiency. Thus, this research aims to (1) develop a novel interpreter modeling system predicated on millimeter-wave radar technology and (2) select the crane as an illustration to investigate the potential applications of this emerging communication paradigm during construction machinery operations. In this investigation, a spatial gesture signal interpreter was devised specifically for machine operators and signalers to augment the quality of communication during the execution of spatial localization tasks. Corresponding limitations that will be encountered in current communication systems were also addressed. This research uses a 60 GHz millimeter wave radar as a gesture trajectory detector, with the benefits of portability and robustness. Its millimeter-level precision enables the capture of highly accurate micro-gestures. The research constructs a novel 3D Spatial Gesture-based Trajectory Modeling system, which will be compared with traditional communication models in future research stages.
Keywords: Communication, Machinery Operation, Gesture Signal Interpreter, Millimeter-wave Radar

1. INTRODUCTION

Heavy construction machinery, including cranes, plays a crucial role in modern construction sites, where the size and complexity of projects have increased. Ensuring productivity and safety in construction projects relies heavily on operating these machinery [1]. However, construction machinery accidents pose significant hazards, with even minor errors or communication issues leading to severe consequences for operators and nearby workers. Effective on-site communication practices are essential for successful task completion. A comprehensive study conducted by Chen et al. [2] examined 95 fatal accidents associated with construction machinery operations in Hong Kong between 2011 and 2021, and their findings identified miscommunication as the primary causal factor of these incidents. Communication issues contribute to operational defects, such as signaler errors, communication failures, inattentiveness, and operator errors, all of which hinder construction progress in one research [3]. A survey conducted by Tam and Fung [4] among crane operators reported that approximately 5.9% of the respondents encountered difficulties comprehending radio communication signals between crew members. One study [5] has highlighted that the process of load handling and communication between operators and signalers has remained largely unchanged for several decades. Consequently, several research endeavors have sought to innovate and transform the system employed by operators and signalers.

In order to address the communication challenges faced by operators and observers, researchers have put forward the proposition of integrating camera systems and real-time video streaming technologies. The primary objective of these solutions is to enhance visibility, augment the situational awareness of operators, and ultimately improve productivity and safety in construction machinery operations. One study [6] investigated the factors influencing workers' adoption of wearable technology in the occupational work environment and found that workers have concerns about perceived privacy risks. Hence, although studies [7] employed camera-based gesture recognition to improve efficiency and quality in the communication between signalers and operators, it is important to acknowledge that privacy concerns may arise among workers. Moreover, the use of cameras may be constrained by factors such as inadequate lighting, blurred lenses, and overexposure. Unlike vision-based systems, a radar sensor can detect and classify targets independent of illumination and operating environment. Recently, radar-based approaches for dynamic hand gesture recognition have attracted much attention from industry and academia [8]. A radar sensor can capture a very small magnitude of motions and can classify very subtle and precise hand gestures. Furthermore, approaches like adopting millimeter-wave radar can safeguard workers' privacy. Previous research [9] has already explored the application of millimeter-wave radar for gesture recognition in automotive driver interfaces. Therefore, we propose utilizing millimeter-wave radar for gesture recognition and hand motion tracking in a one-to-one operator-signalman communication system. The study is to investigate the feasibility of developing a novel communication paradigm in the construction industry using millimeter-wave radar.

2. COMMUNICATION SYSTEMS IN MACHINERY OPERATIONS

This section uses tower cranes to elucidate the progression of the communication system between construction machinery operators and signalers. The construction site environment has witnessed escalating complexities, and the proliferation scale of construction sites become larger and taller, accompanied by increased floor heights. Consequently, the operation of tower cranes has become progressively challenging. In response to these challenges, the communication system between operators and signalers has evolved incrementally to meet the industry's evolving demands. Several facilitating approaches are shown below:

2.1 Hand Signals

The application of hand signals as a prevalent communication method between operators and observers has become prominent with the advent of various construction machinery. International standards, such as the "Crane Hand Signals" released by the International Organization for Standardization (ISO) in 2014 [10], provide a comprehensive set of standardized gestures specific to different crane operations. These signals effectively address the operational requirements of stationary tower cranes, offering instantaneous language-independent communication within the operator's visual range. However, relying solely on hand signals may not sufficiently address the complexities of the construction site environment. Factors such as obstructed line of sight, distance between the signaler and operator, inadequate lighting, and environmental disturbances can hinder the effectiveness of hand signal communication. Consequently, the use of a light-of-sight-based hand signal communication system may impede information exchange promptness and overall operational efficiency. To overcome this limitation, researchers like Cheng et al. [11] have explored methods to enhance operator visibility of ground operations. These studies involve proactive scanning of potential blind spots within the operator's field of vision, accurate identification of nearby workers, and improving the safety of tower crane operations.

2.2 Radio Communication System

In conjunction with hand signals, radio communication serves as the prevailing means of communication in the operation of diverse construction machinery. This communication method often involves a combination of hand signals. For instance, visual contact between the operator and the observers or the load is obstructed, necessitating radio communication to ensure the proper execution of operations. Notwithstanding the convenience and simplicity of radio communication, two key limitations persist: (1) The dedicated channel used by the radio has limitations, and any latency problems can cause delays in operation. When the site is noisy, the operator cannot clearly hear the observers, which may cause operations to be interrupted until the noise problem is resolved; (2) Language problems. People working on construction sites often have different first languages and accents, which can also cause operators to miscalculate while listening to the observers' commands, further causing accidents. All of the above issues can cause operators to make poor decisions, which is not only unsafe but also reduces productivity. In addition, one research [12] has argued that the capacity of traditional information flows limits the shift to a construction paradigm with access to ubiquitous data networks. Therefore, more all-encompassing and effective communication systems are needed.

2.3 Camera and Sensor System

To address communication challenges between operators and observers, scholars have proposed integrating camera systems and real-time video streaming technologies. These solutions aim to enhance visibility, augment operator situational awareness, and improve productivity and safety in crane operations. For example, Everett and Slocum's introduction of the CRANIUM camera monitoring system [13] in 1993 revolutionized communication between operators and signalmen, significantly increasing productivity from approximately 16% to 21%. In addition, researchers have explored the utilization of gesture recognition through cameras and monitors, enabling operators to accurately interpret signals, even when the signaler is not within their visual range. Mansoor et al. [7] proposed a communication framework incorporating image recognition technology, where a camera on the signaler's helmet captures gesture signals, which are then processed and matched with predefined frames stored in a database. This approach allows for the direct acquisition of signaler instructions, enhancing communication even when visual contact is limited.

2.4 Millimeter-wave Radar

Despite being a popular research product in recent years, the application of millimeter-wave radar in construction machinery and construction sites remains less. Currently, millimeter-wave radar primarily finds utilization in domains such as smart furniture and automobile driving assistance systems. Dong et al. [9] introduced a gesture recognition network that leveraged in-vehicle radar, highlighting the advantages of millimeter-wave radar in terms of preserving passenger privacy compared to camera-based gesture recognition methods. Furthermore, millimeter-wave radar offers enhanced accuracy in close-range scenarios and remains unaffected by external factors such as ambient lighting or lens quality. Another notable advantage of millimeter-wave radar is its compact size, exemplified by the Infineon BGT60TR13C radar which we used in this research, as diminutive as a fingernail cap, measuring only 6.5 x 5.0 x 0.9 mm^3.

3. METHODOLOGY

Considering the diversity and complexity of construction machinery, our current research has produced a millimeter-wave radar spatial gesture-based trajectory recognition system using the tower crane's communication system as a case study.

3.1 Operation Gesture Signals Transformation

Considering the range of millimeter-wave radar and the limitation on the confines of scaffolding or cramped spaces, gesture signaling cannot continue using full-body hand signals under the ISO standard. Also, in order to make the signaler's commands more intuitive, we have designed two modes of signal conversion.

One is to directly recognize the gesture direction and convert it into a 2D direction signal. Another one translates the gesture signals directly into 3D motion trajectories. Through the coordinate system transformation, the operator will have a more intuitive understanding of the corresponding operation of the tower crane system. Operators can make their own judgments based on the motion guidance displayed on the computer screen in the cockpit, while at the same time, the signaler can view the results of the commands on the mobile device, helping them to adjust the commands by further gesture signals. Our approach utilizes the three receiving antennas of the radar BGT60TR13C, arranged orthogonally as the spatial coordinate axes. For simplicity, we consider the signaler issues commands with the x-axis aligned parallel to the crane boom and facing the cockpit or operator (as shown in **Fig. 1**), then the direction of motion or spatial vector for the crane would be:

$$\vec{V} = A(\vec{x}, \vec{y}, \vec{z})u$$

The $\vec{x}, \vec{y}, \vec{z}$ are the vectors corresponding to the radar coordinate system, and the A is the scaling factor, which could be adjusted to accommodate the specific dimensions of the crane and the designated working area. u is a 3×3 unit matrix that changes positively or negatively depending on the positional relationship between the radar's and crane's coordinate systems.

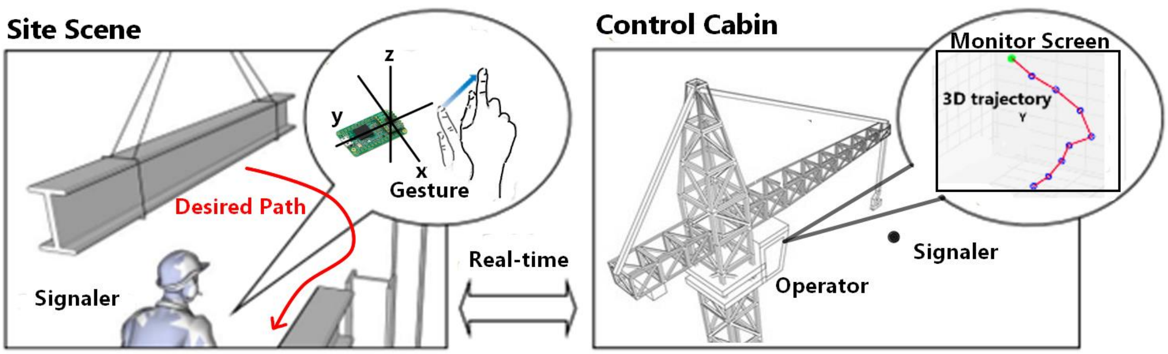

Fig 1. Example of signaler communicating with operator

3.2 Gesture Recognition Model

In this study, we used the millimeter-wave radar BGT60TR13C to capture eight distinct hand gestures, which were subsequently analyzed using Two-Dimensional Fast Fourier Transform (2DFFT) for feature extraction. These hand gestures were specifically designed to correspond to the crane's movements in six directions: forward, backward, left, right, upward, and downward. Furthermore, clockwise and counterclockwise gestures were utilized to initiate 3D trajectories and issue emergency stop commands, respectively. The workflow depicting the experimental process is depicted in **Fig. 2** below:

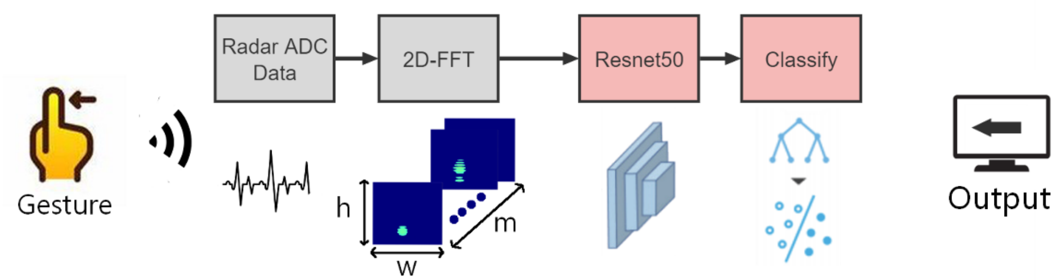

Fig 2. Diagram of the gesture recognition model

This gesture recognition model can be defined as a functional relationship denoted as ϑ.

$$\vartheta: R^{m\times w\times h\times c} \rightarrow R^{n}$$

In this formula, $R^{m\times w\times h\times c}$ represents each input frame to the network, m denotes the number of 2DFFT maps in time series for features (currently set as $m = 6$). w and h represent the width and height of the feature maps, respectively. c denotes the number of channels, which in this context represents the radar echo intensity values, with three radars, hence $c = 3$. For R^{n}, where n represents gesture category, $n = 8$ in this study. The following diagram (as shown in **Fig. 3**) illustrates the performed gesture, the radar-generated signal map associated with it, and the consequent tower crane operation. The presented figure exemplifies the reception of signals from antenna 1. The range doppler and range spectrogram signal images, derived from the captured gesture signals, are utilized in conjunction with a trained model to predict the intended tower crane operation.

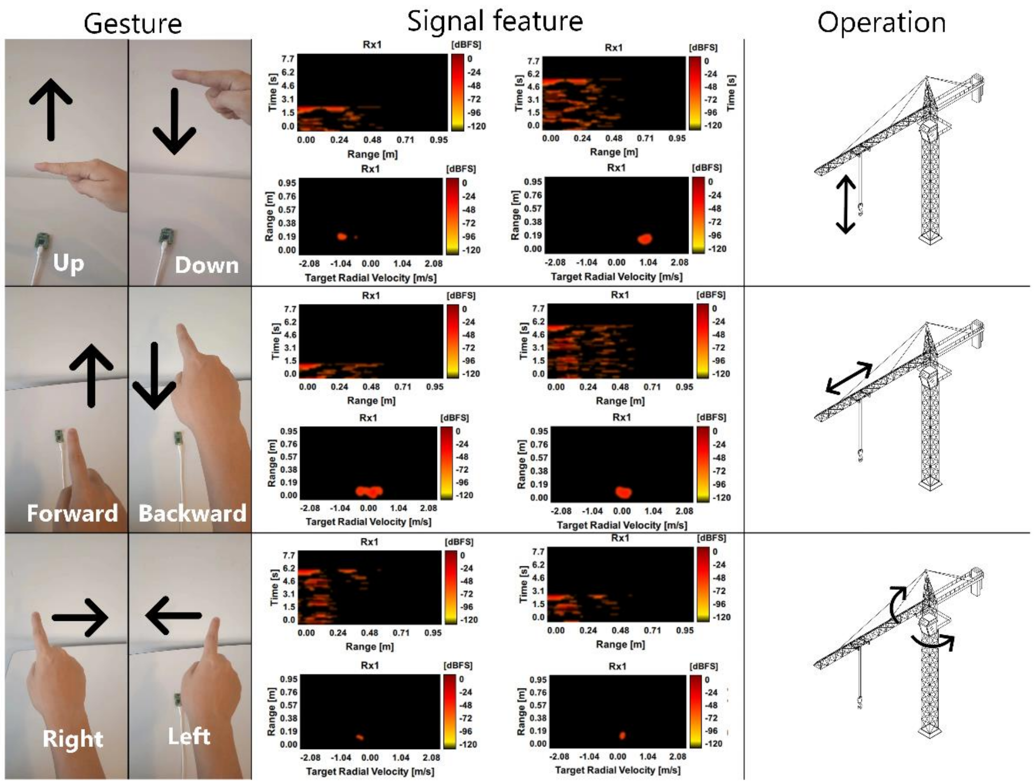

Fig 3. Example of the transformation between gesture and crane operation

3.3 3D Spatial Tracking

To achieve precise 3D trajectory tracking using millimeter-wave radar, a minimum of three receiving antennas is typically required. By employing triangulation techniques, the radar system can accurately determine the object's position and movement in three-dimensional space. In this study, we have chosen the BGT60TR13C radar module, which features three receiving antennas and one orthogonally aligned transmission antenna, ensuring the fundamental tracking capabilities needed for our research. To determine the spatial position of the target object, namely the signaler's hand, within the coordinate system established by the millimeter-wave radar, it is crucial to ascertain the distance between the receiving antenna and the target object, as well as the pinch angle. The distance measurement process for Frequency-Modulated Continuous Wave (FMCW) radar is relatively straightforward, and the following formulas can be utilized:

$$d = t_d c/2$$

In distance measurement, t_d is the time interval between the transmission of electromagnetic waves from the transmitting antenna and the reception of electromagnetic waves by the receiving antenna. By analyzing the phase change and distance variation, the following formula can be derived for estimating the arrival angle:

$$\begin{cases} \theta = \sin^{-1}\left(\frac{\lambda \Delta\phi}{2\pi l}\right) \\ \Delta\phi = \frac{2\pi \Delta d}{\lambda} \end{cases}$$

In which Δd is the difference in distance, l is the distance between two receiving antennas, λ is the wavelength. In this formula, $\Delta\phi$ is dependent on the sine function and exhibits a nonlinear relationship. Consequently, the angle estimation accuracy is highest at 0° and diminishes as the angle approaches 90°. For our system, the angle range has been set to 45°. The angle θ_y in the y-axis can be computed using the values from receiving antennas 1 and 3 (Rx1 & Rx3), while the angle θ_z in the z-axis can be determined by considering receiving antennas 2 and 3 (Rx2 & Rx3). Additionally, the angle θ_x in the x-axis can be obtained by calculating the cosine of the squared vector angles. Consequently, the coordinates of the current object within the spatial coordinate system can be expressed as follows:

$$V_t = (d_t cos\theta_x, d_t cos\theta_y, d_t cos\theta_z)$$

Due to the utilization of millimeter-wave radar operating at frequencies of 60 GHz, the distance error incurred by the target object during the reception of electromagnetic waves is essentially negligible. Consequently, by employing calculations involving both angle and distance, the spatial coordinate records of the target object within the coordinate system can be obtained, and the following matrix set can be obtained through multiple records:

$$\overrightarrow{V_{t_0 \sim t_n}} = (V_{t_0}, V_{t_1}, V_{t_2}, \ldots, V_{t_n})$$

In order to mitigate potential errors and prevent unintended tracking of unrelated objects, we used the defined gesture types to perform point trajectory recognition of the hand simultaneously. The system starts recording 3D positions only upon successfully recognizing a clockwise rotation gesture performed by the hand. Currently, the system is configured to record the 3D position at intervals of 0.25 seconds, resulting in a total of nine sets of location points over a 2-second duration. These points are connected sequentially to generate a comprehensive 3D trajectory view. The green point (as shown in **Fig. 4**) is the start point at the time t_0 after the clockwise rotation of the recognized gesture. It is worth noting that the recording duration and frequency can be adjusted to align with specific requirements dictated by the operational context.

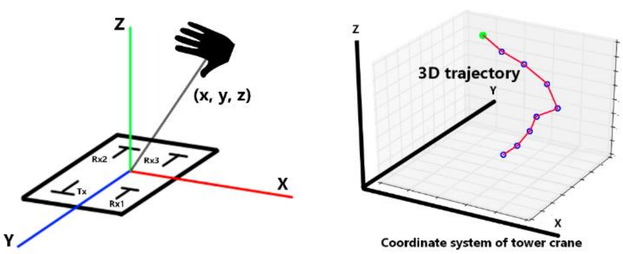

Fig 4. Schematic diagram of spatial tracking (left) and plot trajectories (right)

4. RESULTS AND DISCUSSIONS

4.1 2D Gesture Recognition Training Results

In this study, we opted to extract and train 2DFFT images for eight distinct gestures, including up, down, left, right, backward, clockwise, and counterclockwise rotation. After collecting the data samples, the designed eight gestures are extracted, and 1200 samples for each gesture are extracted as the dataset, which is divided into a training set and test set in a 5:1 ratio. The network algorithm framework is based on Tensorflow 2.15; the batch size is 32; the initial learning rate is 0.001, and the training epoch is 50/100. The hardware parameters are: GPU: NVIDIA GeForce RTX3060-27.9GB; CPU: Intel Core i7-12700@2.1GHz, and Memory: 32GB.

Table 1. The gesture recognition accuracy with different models

Model	Accuracy in 50 epochs	Accuracy in 100 epochs
Vgg19	0.5312	0.3742
Resnet50	**0.9524**	**0.9704**
EfficientNet	0.8565	0.8939

Through the test, we observed that employing the Resnet50 model got the most favorable results, achieving an accuracy of 97% after training for 100 epochs (as shown in **Table 1**).

4.2 The 3D Spatial Tracking

For the 3D spatial trajectory tracking results, as a preliminary observation, we conducted a subjective evaluation based on the level of matching between the generated trajectories shown on the screen display and their corresponding gesture waving performed by signalers. Taking into account the previously mentioned 3D spatial trajectory tracking by clockwise start, combined with the previously tested gesture recognition accuracy, the accuracy of 3D spatial trajectory tracking is acceptable, showing the feasibility of interpreting the millimeter-wave radars' signals in determining potential construction machinery operations. The generated trajectories basically matched the desired operation in 20 testing attempts. The following series of images (as shown in **Fig. 5**) show the frames corresponding to this gesture trajectory in the video recording, and it can be found that the radar-recorded trajectory basically agrees with the actual running trajectory.

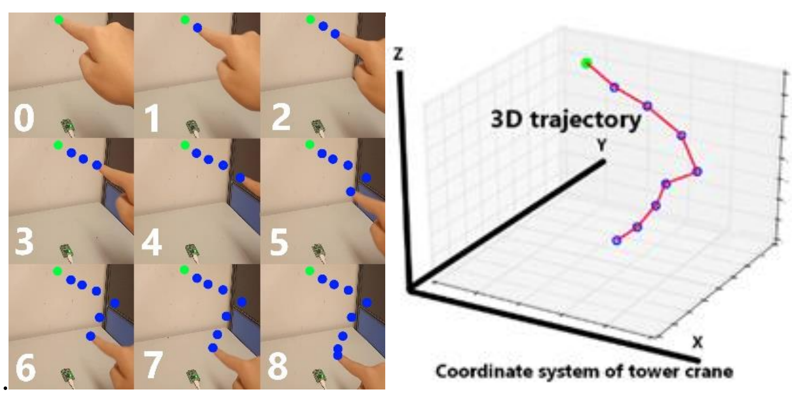

Fig 5. Hand trajectory frames and trajectory tracking result

4.3. Existing Problems and Limitations

The preliminary research shows the feasibility of using millimeter-wave radar signals for gesture recognition during operator-signaler communication of crane operations. Further problems and limitations that need further study are discussed as follows. Firstly, as discussed in the preceding section, the current experiment lacks precise measurements for accurately recognizing the trajectory of the signaler's hand in 3D space. They are also subject to signalers' subjective judgment in interpreting spatial operation tasks. The current testing methodology only provides a general overview of the trajectory without accounting for detailed deviations, thus requiring future work for follow-up. Secondly, the system has not been subjected to experimentation within a real-world building operating environment, nor has it been compared against alternative communication and exchange systems. Consequently, the practical applicability and value of the system remain to be determined. Furthermore, it is important to note that the gesture data used in this study is derived from a limited sample of hand gestures. The model's predictions may be influenced to varying degrees when encountering different individuals' hand shapes, particularly considering the use of gloves commonly worn on construction sites. This factor could potentially impact the accuracy of the model's predictions and will need further investigation.

ACKNOWLEDGMENTS

The authors would like to thank the Research Grants Council, Hong Kong, for their funding support under the General Research Fund (PolyU 25221519).

REFERENCES

[1] Bernold E. Leonhard, Lorenc J. Steven and Luces Erick (1997). "On-Line Assistance for Crane Operators" *Journal of Computing in Civil Engineering*, 11(4),248-259.

[2] Chen, J., Chi, H. L., Du, Q., & Wu, P. (2022). "Investigation of Operational Concerns of Construction Crane Operators: An Approach Integrating Factor Clustering and Prioritization". *Journal of Management in Engineering*, 38(4), Article 04022020.

[3] Raviv, G., Fishbain, B., & Shapira, A. (2017). "Analyzing Risk Factors in Crane-related Near-miss and Accident Reports", *Safety Science*, 91, 192-205.

[4] Tam, W. Y., & Fung, W. H. (2011). "Tower Crane Safety in the Construction Industry: A Hong Kong Study", *Safety Science*, 49(2), 208–215.

[5] Zekavat, Payam & Bernold, Leonhard. (2014). "Embedded Wireless Communication Platform Addresses Crane Safety and Efficiency". 309-318. 10.1061/9780784413517.032.

[6] Choi, B., Hwang, S., & Lee, S. (2017). "What drives construction workers' acceptance of wearable technologies in the workplace?: Indoor localization and wearable health devices for occupational safety and health". *Automation in Construction*, 84, 31-41.

[7] Mansoor, A., Liu, S., Ali, G.M., Bouferguene, A., & Al-Hussein, M. (2020). "Conceptual Framework for Safety Improvement in Mobile Cranes".

[8] K. A. Smith, C. Csech, D. Murdoch and G. Shaker, "Gesture recognition using mm-wave sensor for human-car interface", IEEE Sens. Lett., vol. 2, no. 2, pp. 1-4, Jun. 2018.

[9] DONG Lianfei, Ma Zhixiong, ZHU Xichan. "Dynamic Gesture Recognition Network Based on Vehicular Millimeter Wave Radar[J]". Transactions of Beijing institute of Technology, 2023, 43(5):493-498.

[10] ISO 16715:2014 "Cranes — Hand signals used with cranes"

[11] Cheng, Tao & Teizer, Jochen. (2013). "Modeling Tower Crane Operator Visibility to Minimize the Risk of Limited Situational Awareness". *Journal of Computing in Civil Engineering*. 28. 10.1061/(ASCE)CP.1943-5487.0000282.

[12] Zekavat, P. R., & Bernold, L. (2014). "Embedded Wireless communication platform addresses crane safety and efficiency". *In Construction Research Congress 2014: Construction in a Global Network* (pp. 309-318).

[13] Everett J G, Slocum A H. "CRANIUM: Device for improving crane productivity and safety[J]". *Journal of construction engineering and management*, 1993, 119(1): 23-39.