# Effects of Visual Prompts in Human-machine Interface for Construction Teleoperation System

**Yeon Chae[1], Samraat Gupta[2] and Youngjib Ham[3]**

[1]Department of Construction Science, Texas AandM University, USA
[2]Department of Computer Science and Engineering, Texas AandM University, USA
[3]History Maker Homes Endowed Associate Professor, Department of Construction Science Texas AandM University, USA

yeonchae62@tamu.edu, samraatg@tamu.edu, yham@tamu.edu

**Abstract**

**In construction teleoperation, particularly in disaster restoration, delicate manipulation of heavy machinery is crucial, based on a thorough understanding of the surroundings. Current practices have utilized multiple viewpoints to facilitate a thorough understanding of the site's 3D spatial layout. However, challenges might arise as visual cues within the surroundings could create distractions for teleoperators. Drawing from visual search theory and Gibson's perception theory, exploring visual prompts in teleoperation interface could enhance performance by directing attention to key visual cues, reducing cognitive workload. Nonetheless, the evaluation of different visual prompts from human factors perspectives has been underexplored. Addressing challenges of potential distraction from multiple viewpoints and inappropriate visual prompts, this study emphasizes the necessity of exploring different visual prompts to most effectively guide operators' attention in given work environments. The experiment, designed with low and high visual cue environments, focuses on debris removal and extended 3D Fitts' law tasks, evaluating spatial awareness and depth perception during teleoperation. The experiments were conducted with participants in construction-related fields with industrial experience. Performance measurements and subjective ratings with open-ended discussion was conducted. The findings show visual prompts' effects on distraction and visibility conditions concerning task-oriented difficulty levels in teleoperation. The experimental results can inform the optimal design of visual prompts in human-machine interface for teleoperation for complicated construction environments, highlighting the importance of the considerations of environments and task characteristics.**

**Keywords –**

**Human-Machine Interface, Construction Teleoperation**

## 1 Introduction

In the aftermath of disasters and demolition, hazardous conditions within sites create barriers to physical human access [1]. In such cases, teleoperation technology has been adopted, enabling workers to operate heavy machinery from a secure external location, minimizing on-site risks [2]. Despite this advancement, such jobsites present persistent challenges in teleoperation that directly hinder productivity, especially in constrained movement areas where teleoperators require heightened spatial awareness [3]. To alleviate these challenges, incorporating multiple viewpoints to offer additional information seems promising, yet the potential overflow of information might overwhelm teleoperators [4,5]. Previous studies have highlighted the efficacy of visual prompts in guiding teleoperators to conduct critical tasks with multiple screen settings [4]. However, the exclusive efficacy of visual prompts in teleoperation and their alignment with the naturalness principle remains unexplored, along with an incomplete understanding of how environmental settings influence the effects of visual prompts in teleoperation.

This paper aims to explore varied types of visual prompts, by differentiating the forms of conveying spatial information and the level of granularity in indicating distances to the target objects, to uncover their effectiveness in teleoperation. Examining these conditions offers refined insights into how different visual prompts impact teleoperators' performance, and task difficulty and efficiency. This study can inform how to optimize visual interfaces for teleoperation in different task environments and address potential challenges of visual prompts in certain environment settings. Through the investigation of varying impacts of different visual prompts on teleoperation across diverse task settings, this study addresses critical gaps in understanding and contributes to the visual interface design in teleoperation systems, especially heavy machinery. Overall, this study can provide practical implications for enhancing teleoperator experience, refining task outcomes, and advancing human-robot interaction.

# 2 Backgrounds

## 2.1 Visual Prompts in Human-Machine Interfaces for Teleportation

The visual interface for human-machine teleoperation heavily relies on understanding selective attention through visual search theory and Gibson's Theory of Perception [6]. Teleoperators engaged in articulate manipulation tasks necessitate sustained attention, making the design of a robust visual interface crucial [5]. Visual prompts, guided by top-down and bottom-up influences, affecting the adjustment of bottom-up approach by the salient features and top-down approach by controlling themselves how to utilize the given visual information based on their goal, also play a crucial role in human-robot interaction, directing attention toward essential elements [4]. The design of these prompts should reduce involuntary attention shifts by carefully considering their reliability, following Gibson's ecological approach, which highlights the importance of perceiving environmental cues that support the robot's actions [7]. Understanding selective attention mechanisms and incorporating visual search theory principles can markedly enhance visual prompts' design, optimizing teleoperator performance and improving human-robot interaction during teleoperation [4].

Aided visual prompts in robotic teleoperation encompasses a range of types, addressing specific environmental and task needs, facilitating tailored prompts that enhance operators' situational awareness [8]. The selection of optimal visual stimuli can significantly influence operators' decision-making and overall task performance, aligning with the specific teleoperation requirements [9]. In the realm of teleoperation, visual prompts categorize into directional and regional types [9], each conveying distinct information. Direct visual prompts, particularly regional ones, aim to provide 3D spatial understanding, either directly or indirectly, crucial in scenarios where desktop displays offer 2D vision. These prompts aim to engage voluntary attention, improve goal-directed task execution, and reduce cognitive demands, significantly impacting teleoperation performance [4]. Strategic implementation of these visual prompts, emphasizing relevant features in alignment with top-down approaches, optimizes operators' attentional resources, ultimately enhancing teleoperation efficiency [4,10].

## 2.2 Teleoperation Task Environments with Varying Level of Visual Cues

Level of visual cues could be varied by the extent to which information is provided in terms of depth perception [12]. Based on the hierarchy of the visual cues [13], there is a property of visual cues that could assist the depth perception. Based on the experiment conducted [14], low level of visual cues provides minimum spatial cues (including relative size of the objects, occlusion, linear perspective) whereas high level of visual cues provides maximum spatial cues (including distance to horizon with familiar visual cues, cast shadows).

Depth perception in varied visual environments, characterized by cue availability, poses significant challenges in tasks, especially when working with low visual cues lacking familiar landmarks [15]. Evaluating cue reliability becomes critical as the inclusion of unreliable cues can hinder rather than aiding task performance [16]. Ensuring cue accuracy and reliability is crucial to prevent misinterpretations in scenarios where depth perception uncertainties are amplified due to the absence of adequate visual cues [16]. However, in the context of teleoperation, these challenges might be exaggerated as the task itself has to be manipulated in the complicated environments which require a delicate maneuver, engaged by continuous cognitive engagement of teleoperator. Moreover, as given that the teleoperator is provided by the visual interface, indirectly experience the scene through it, there’s a room to augment that interface, by varying the level of visual cues given to the teleoperator with a thorough consideration of the task characteristics and environments. Therefore, integrating multiple visual cues into the interface aims to enhance performance and usability, necessitating a keen understanding of cue reliability's impact on decision-making and task execution [10].

# 3 Modeling

The overview of this study is depicted in Figure 1. We build upon the two models with different visual cues, a low visual cue environment as a baseline setting and the demolition sites as a high visual cue environment.

## 3.1 Task Environment with Low-Level Visual Cues

The virtual space featured a clear blue-sky area strategically designed to minimize distractions for participants [17]. In our study, the low-level visual cue environments were built upon the Fitts' law task [17],


Figure 1. Overview of this study to answer "How can we most effectively map spatial data to enhance situational awareness?" by exploring diverse representations of visual prompts and task attributes in construction teleoperation.

emphasizing visual cues limited to target, destination, and a horizontal line, serving as a key perceptual reference [13]. These cues, influencing depth perception, encompass occlusion, relative size, distance to the horizon, and linear perspective [13]. Given the pivotal role of target size and distance in task complexity [17], the target consistently commences midway between center and destination. Each trial situates the destination around the target with identical 2D distances but random forward or backward depths in 3D, maintaining uniformity in depth radius.

The design units of objects and camera length aim for comprehensive scene understanding across varied camera conditions. The scenario introduces the destination in horizontal, vertical, and diagonal locations, each presented twice, totaling six occurrences. In each of these locations, both forward and backward depths are assigned randomly per trial, although the sequence is randomized, ensuring consistent difficulty across manipulation trials.

### 3.2 Task Environment with High-Level Visual Cues

In this experiment, the demolition site is designed as an environment with high-level visual cues. This included a demolished building and an excavator. The excavator mirrored real-world behavior, and the trajectory of excavator is controlled for the safety measures to prevent tipping on slopes during the manipulation. Camera placements were tailored for cabin-installed position. In this environment, unsound structures are near excavators, which requires careful operation. Dust effects and sound cues when collision occurs as collision effects are designed to emphasize immersion. As illustrated in Figure 2(a), debris are occluded, invisible from the cabin view. Following the task design, the environment is designed to restrict the excavator's trajectory with the closely located obstacles (Figure 2(b)), with the dumping area (Figure 2(c)).


Figure 2. Environment setting – (a): Debris, (b): Overall View of the Scene, (c): Dumping Area.


Figure 3. Visual prompts design in demolition sites – (a): Continuous representation, (b): Discrete

## 3.3 Visual Prompts Design

Basically, visual prompts can serve to enhance comprehension of 3D spatial information in remote construction sites [4]. In our experiments, components used in the model are selectively adopted from previous studies [4,10] and customized to the given experimental scenarios and task settings. The primary aim of these visual prompts is to augment understanding, especially regarding depth perception within a 2D display representing 3D spatial information. These prompts encompass line interpolation, distance, and color coding, assisting in removing debris and monitoring progress within occluded areas. Categorized as work facilitation and obstacle avoidance prompts [4], these prompts aim to support situational awareness for the task including the debris removal and obstacle avoidance.

For "work facilitation prompts" aiding debris removal task, visual cues include the productivity monitor bar conveying the information of the number of debris being completed in picking up and dumping task, green lines guiding the nearest debris mass center, and proximity-indicating lines (Figure 3). These lines adapt color and thickness based on environmental settings, providing intuitive cues for obstacle proximity. In "obstacle avoidance prompts", lines indicate the closest vertical points of obstacles and excavator components, varying in width to represent depth. Continuous prompts display numerical distance values and color gradients indicating proximity (Figure 3(a)). Discrete prompts feature warning signs triggered by distance thresholds, with colors indicating danger levels (Figure 3(b)). In discrete visual prompts, the distinction between "warning" and "danger" levels is determined by proximity criteria [4]. Similarly, work facilitation prompts categorize proximity as "close" or "very close," based on the distance relative to the stick's length. Contrastingly, continuous visual prompts use numerical values in feet, detailed to four decimal points, to indicate proximity. When the participants operate the excavator, the continuous visual prompt changes occur at a frequency of 60 fps. However, the discrete visual prompt remains unchanged while a part of the machine is physically within the predefined range. In the baseline scene, visual prompts primarily reflect features from demolition sites, focusing on work facilitation without collision avoidance requirements. These prompts include varied line attributes indicating depth in both discrete and

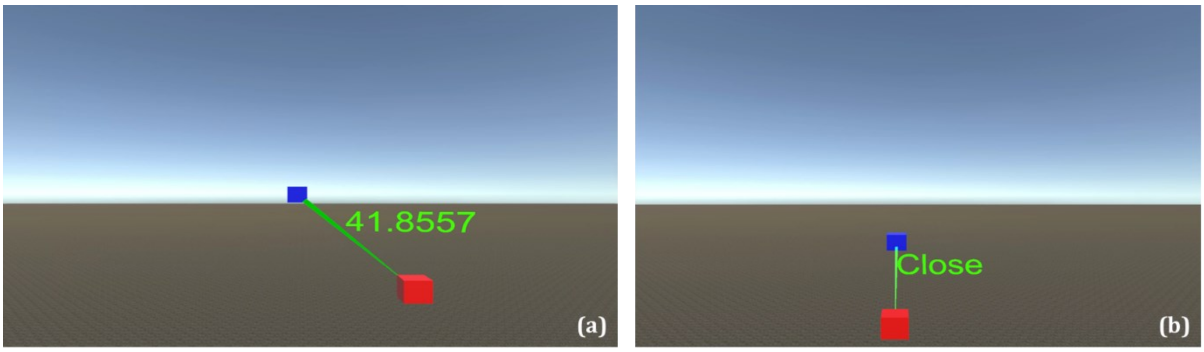


Figure 4. Visual prompts design in low level visual cues environment – (a): Continuous representation, (b): Discrete representation.

continuous settings (Figure 4).

In our experiments, we explore how different visual prompts work in teleoperation for different working environments and look into their effects from human factors perspectives and how to most effectively map spatial data to enhance situational awareness.

# 4 Experiments

## 4.1 Task Design

In the environment with the high-level visual cues, the task focuses on assessing spatial awareness and depth perception in debris removal scenarios like disaster restoration or demolition sites. Participants maneuver an excavator to pick up debris, avoid obstacles, dump debris, and return two trials within a six-minute time limit, assessing productivity and safety.

In the environment with the low level of visual cues, the extended 3D Fitts' law task evaluates 3D spatial awareness by manipulating objects in horizontal, vertical, and diagonal directions. This task requires six sequential movements of the target cube (indicating as blue) toward the destination where the red cube is located. Destination disappears once each hit is completed and pops up in a random position. The task involves six hits, each with a random position, with two locations set diagonally, vertically, or horizontally. Fixed target sizes and distances provide consistent difficulty levels across movements.

## 4.2 Performance Metrics

In debris removal task, task performance was measured as an objective measure in terms of amount of work done, time and collision occurrences [2]. For extending 3D Fitts’ law task, performance was measured based on time, errors, and efficiency in navigating the movements in a virtual space [17]. Considering the task characteristics, if the participant mistakenly manipulates the joystick resulting in deviating from the target destination, it is counted as an error in this task. This extension of the 3D Fitts' law task delves into the challenges of navigating a given environment in X, Y, and Z directions, considering factors like movement time, error rates, and directional complexities. Understanding these challenges aids in enhancing 3D interfaces for teleoperation, especially concerning different movement difficulties and their impact on task performance.

## 4.3 Effects of Visual Prompts in Different Task Environment Settings

The experiments started with a brief introduction of the study, followed by the task description detailed in section 4.1. Participants then completed tasks, including the extended 3D Fitts’ law task and debris removal task in demolition sites, in a random order. Ten undergraduate students majoring in construction-related fields with industrial experience from Texas AandM University participated. Objective measures were collected during the experiments, along with subjective ratings on task difficulty [4] via a comprehensive questionnaire and open-ended discussions.

Table 1. Pre-survey results for the participants (*N*=10)

| Variables | Scale | # |
|---|---|---|
| Age | 19~21 | 8 |
| | 22~24 | 2 |
| Gender | Male | 8 |
| | Female | 2 |
| VR Experience | Very unfamiliar | 2 |
| | Unfamiliar | 3 |
| | Somewhat familiar | 3 |
| | Familiar | 1 |
| | Very familiar | 1 |

For the extended 3D Fitts’ law task, the completion

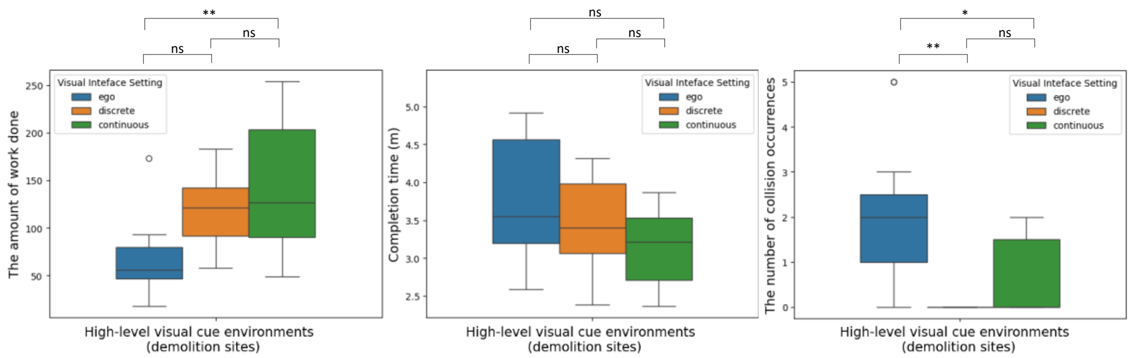


Figure 5. Results from debris removal task in demolition sites with high-level visual cues - (a): Amount of work done, (b): Consumed time, (c): Collision occurrences.

time for each trial (totaling six completion times for a single task) is documented, alongside the coordinates of the blue box (target) and red box (destination). These coordinates provide essential data for tracking the real-time distance between the target and destination throughout the manipulation process. For demolition sites, completion time, amount of work done, collision occurrences are used as performance metrics.

## 5 Results and Discussion

With the results, a comparative examination of group disparities is performed through One-way ANOVA (ANalysis Of VAriance) with post-hoc Tukey HSD (Honestly Significant Difference) Test.

### 5.1 Effects of Visual Prompts in Different Task Environment Settings

Figure 5 shows the performance in demolition sites, as an objective measure in terms of amount of work done, time and collision occurrences. Overall, visual prompts enhance the work performance across all measures compared to the counterpart with non-visual prompts. Interestingly, in terms of collision, comparing the average value, when visual prompts are provided in a discrete way, collision reduced, which is significant compared to the counterpart with non-visual prompts. Especially for continuous visual prompts, based on the open-ended discussion following the experiment, participants replied that they felt exhausted in tracking the continuously changing number, finding it challenging to gauge its proximity and its potential impact on collision avoidance. This lack of clarity made planning maneuvers with heavy machinery difficult, resulting in the occurrence of collisions. Based on the given information (i.e., number), it is hard to plan the maneuver of heavy machinery, since the number itself was not intuitive to inform the control based on that.

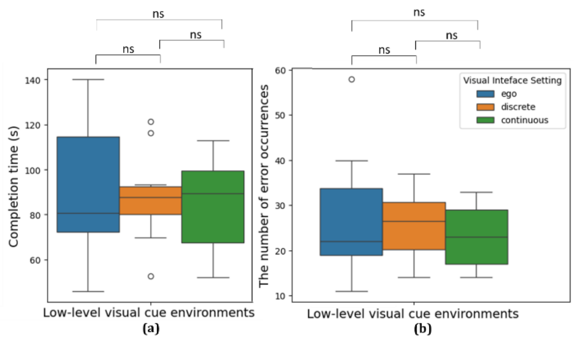


Figure 6. Results in extended 3D Fitts' law task - (a): Consumed time, (b): Error.

Contrarily, when visual prompts are provided with discrete representation, participants reported that their focus was easily directed by the visual prompts, leading to a significant reduction in collision. These varied patterns are also aligned with their interviews, demonstrating that continuous visual prompts in the environments with high-level visual cues could potentially serve as a distraction, impacting teleoperators' attention.

On the contrary, in the low-level visual cues in Figure 6, the average value of the consumed time with visual prompt has increased compared to non-visual prompt conditions. In the environments with low level of visual cues, some participants experienced improved performance with visual prompts, while others encountered hindrances in their manipulative tasks, resulting non-significant difference within different settings ($p > 0.05$). This variance underscores the complexity of how visual prompts influence teleoperators' task execution in low-cue settings, where the impact on performance is not uniformly beneficial and could vary based on individual perceptions and adaptability to visual cues. Interestingly, despite these varied objective outcomes, participants reported lower task difficulty levels when visual prompts were available. Comparing the average number of errors occurrences between two different conditions of visual prompts, continuous visual prompts condition shows lower error incurred, compared to the discrete visual prompts. This result could be originated from the task requirements, as the task requires continuous monitoring of comparing the position of the object being manipulated in relation to its intended destination. Overall, the results highlight the criticality of the visual representation following the naturalness principle based on the consideration of task characteristics within the context of environments.

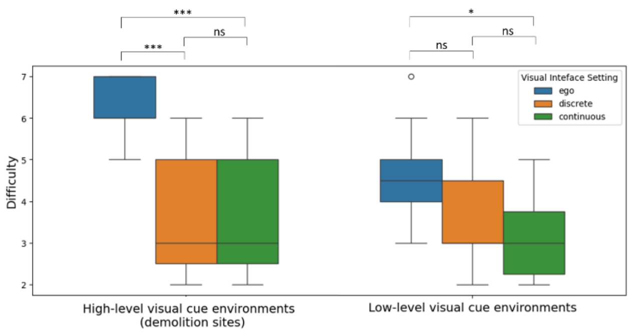


Figure 7. Difficulty level of task for debris removal task and extended 3D Fitts' law task.

Based on the level of visual cues, the results, as illustrated in Figure 7, indicate a significant disparity in perceived task difficulty among participants. In low-level visual cue environments, participants reported feeling less difficulty when engaged in the extended 3D Fitts' law task compared to the high-level visual cue environment where the task involved debris removal.

This suggests that the complexity of the task itself, coupled with the amount of visual information available, plays a crucial role in how participants perceive and experience task difficulty.

These results based on their subjective difficulty level of the task could underscore the necessity of tailoring the interface and visual prompt design based on the environmental context. In low-level visual cue settings, where tasks might be less visually demanding, simpler, and more direct visual prompts could suffice. Conversely, in high-level visual cue environments where tasks are inherently more complex due to the visual richness, visual prompts should offer more comprehensive and easily interpretable information to assist users effectively. The results could highlight that visual prompt designs should adapt to the environmental setting, aligning with task complexities and varying levels of visual cues to optimize user performance and minimize task difficulty.

### 5.2 Effects of Visibility Conditions and Task-Specific Difficulty Levels

Participants replied that highlighted challenges were linked to both discrete and continuous signals with the visibility condition of the working area, particularly in discerning object proximity during their task. In our scene, during the picking up task, the target object is occluded, whereas during the obstacle avoidance task, the obstacle, the target to avoid, is visible, not occluded. Specifically, discerning the proximity of objects, notably when occluded, posed significant difficulties for teleoperators.

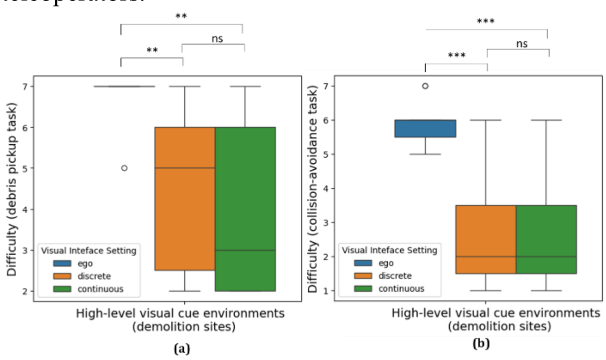


Figure 8. Difficulty level in debris removal task – (a): Picking up, (b): Collision avoidance.

During the debris picking up task, with the condition of discrete representation of visual prompts, defining the distinctions between being "close" or "very close" became ambiguous during manipulations. The sign was too vague to grasp the relative proximity between the target object (debris) and the end effector of the excavator (bucket). Contrarily, the condition of continuous representation of visual prompts, offered comprehensive proximity details, teleoperators encountered issues in ascertaining movement orientation beforehand, requiring them excessive manipulation to check which direction they should input.

The results indicate the effect of visibility conditions to task difficulty, considering work environments as a key factor which could directly affect the teleoperator's performance and task complexity, as shown in Figure 8. In these high cue environments, the abundance of data could overwhelm operators, underscoring the crucial need for intuitive data presentation. For instance, discrete signals are typically employed to keep workers alert, while continuous representations offer accurate depth perception but can be overwhelming. Interestingly, teleoperation tasks in the high-level visual cue environments were more challenging than the low-level visual cue environments, with low visual cues posing challenges due to the absence of visual cues to infer depth more accurately. These challenges encompass diverse factors such as joystick manipulation, task intricacies, and the varying levels of visual cues.

This study endeavors to unravel the complex interactions among task performance, environmental context, and the impact of visual prompts, aiming to uncover the challenges involved. This investigation illuminates the varied challenges faced by teleoperators concerning different types of visual prompts within different levels of visual cue environments.

## Conclusion

Visual interface in human-machine teleoperation is crucial particularly in addressing selective attention through visual search theory and Gibson's Theory of Perception. This paper demonstrates the crucial role of customized visual interfaces in enhancing human-robot interaction by guiding operators' attention toward essential task elements while reducing involuntary attention shifts. The investigation into depth perception complexities in varied visual environments emphasizes the necessity of assessing cue reliability to prevent performance hindrances. The comparative analysis between non-visual and visual prompt settings demonstrates the varied impacts of different visual prompts on teleoperators' performance, cognitive load, and task efficiency across diverse environments and tasks. This highlights the criticality of visual representation formats in influencing performance, with varying the environments with different level of visual cues where the effects of visual prompts on task execution vary based on individual perceptions. Moreover, this study emphasizes the need for further exploration of visual prompts' effects concerning diverse environmental and task settings, shedding light on the intricate interaction between teleoperator performance, task characteristics, and environmental factors. The investigation provides insights into the dynamics between visibility conditions and task-oriented difficulty levels, presenting challenges and implications for future studies in optimizing visual interfaces for teleoperation, especially in contexts where heavy machinery is used.

## References


[1] Shigematsu K., Tsubouchi T. and Sarata S. Tip-over prevention system based on motion prediction for teleoperated excavator. *Advanced Robotics*, 35, 1438–1449, 2021.

[2] Liu D., Kim J. and Ham Y. Multi-user immersive environment for excavator teleoperation in construction. *Automation in Construction*, 156, 105143, 2023.

[3] Kamezaki M., Yang J., Iwata H. and Sugano S. Visibility Enhancement using Autonomous Multicamera Controls with Situational Role Assignment for Teleoperated Work Machines. *Journal of Field Robotics*, 33, 802–824, 2016.

[4] Kamezaki M., Yang J., Sato R., Iwata H. and Sugano S. A situational understanding enhancer based on augmented visual prompts for teleoperation using a multi-monitor system. *Automation in Construction*, 131, 103893, 2021.

[5] Motohashi S., Qiao Z., Moteki T. and Iwata H. Analysis of the Effect of the Operator's Spatial Cognition on Their Planning Skills in Unmanned Construction. In *2023 IEEE/SICE International Symposium on System Integration (SII)* (pp. 1–7), 2023.

[6] Müller H. and Krummenacher J. Visual search and selective attention. *Visual Cognition*, 14, 389–410, 2006.

[7] Hong Z., Zhang Q., Su X. and Zhang H. Effect of virtual annotation on performance of construction equipment teleoperation under adverse visual conditions. *Automation in Construction*, 118, 103296, 2020.

[8] Li X., Fan J. and Su X. Influence of Spatial Ability on Virtual Annotation Response in Construction Equipment Teleoperation. In J. Li, W. Lu, Y. Peng, H. Yuan, and D. Wang (Eds.), *Proceedings of the 27th International Symposium on Advancement of Construction Management and Real Estate* (pp. 776–785). Springer Nature, 2023.

[9] Livatino S., Guastella D., Muscato G., Rinaldi V., Cantelli L., Melita C., Caniglia A., Mazza R. and Padula G. Intuitive robot teleoperation through multi-sensor informed mixed reality visual aids. *IEEE Access*, 9, 25795-25808, 2021.

[10] Arevalo Arboleda S., Rücker F., Dierks T. and Gerken J. Assisting Manipulation and Grasping in Robot Teleoperation with Augmented Reality Visual Cues. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–14). Association for Computing Machinery, 2021.

[11] Wang M., Wong P., Luo H., Kumar S., Delhi V. and Cheng J. Predicting Safety Hazards Among Construction Workers and Equipment Using Computer Vision and Deep Learning Techniques. In ISARC. *Proceedings of the International Symposium on Automation and Robotics in Construction* (Vol. 36, pp. 399-406). IAARC Publications, 2019.

[12] Burge J., Fowlkes C. and Banks M. Natural-scene statistics predict how the figure–ground cue of convexity affects human depth perception. *Journal of Neuroscience*, 30(21), 7269-7280, 2010.

[13] Bogdanova R., Boulanger P. and Zheng B. Depth Perception of Surgeons in Minimally Invasive Surgery. *Surgical innovation*, 23(5), 515-524, 2016.

[14] Park H., Faghihi N., Dixit M., Vaid J. and McNamara A. Judgments of Object Size and Distance across Different Virtual Reality Environments: A Preliminary Study. *Applied Sciences*, 11, 11510, 2021.

[15] Newman P. and McNamara T. Integration of visual landmark cues in spatial memory. *Psychological Research*, 86, 1636–1654, 2022.

[16] Jacobs R. What determines visual cue reliability? *Trends in Cognitive Sciences*, 6, 345–350, 2002.

[17] Clark L., Bhagat A. and Riggs S. Extending Fitts' law in three-dimensional virtual environments with current low-cost virtual reality technology. *International Journal of Human-Computer Studies*, 139, 102413, 2020.