A Study on the Construction of a Visual Presentation Method That Can Prevent Cognitive Tunneling in Unmanned Construction

Takumi Moteki, Ziwei Qiao, Yuichi Mizukoshi, and Hiroyasu Iwata

Department of Creative Science and Engineering, Waseda University, Japan
E-mail: motekitakumi@gmail.com, qiaoziwei1999@gmail.com, yuichi_9432@akane.waseda.jp, jubi@waseda.jp

Abstract–

One of the problems with unmanned construction is the lack of visual information, which reduces work efficiency to less than half of that in onboard operation. Therefore, methods to provide visual information using drones and image processing were studied in the past. However, the addition of information causes the operator to fall into cognitive tunneling in which the attention is focused only on a specific image. In this study, we attempted to develop a method that can prevent cognitive tunneling and shift the operator attention to an appropriate view according to the working state of heavy machinery. Cognitive tunneling is caused by low visual momentum (which represents ease in information integration between views) and high visual saliency (which represents ease in attention). Therefore, because visual momentum can be improved by presenting the same landmark in different camera images, useful landmarks for each work state were included in each image. In addition, because humans tend to pay attention to objects that vibrate in the useful field of view, we presented the image of an external camera in the useful field of view and allowed the image to vibrate when the work state was switched. To investigate the effectiveness of the proposed method, an experiment was conducted on an actual hydraulic excavator. Although the proposed method did not improve the work efficiency of some operators, we believed that the proposed interface could direct the eyes of the operator to an appropriate image according to the work state.

Keywords–

Unmanned construction; Remote operation; Cognitive tunneling; Visual momentum; Visual saliency; Visual support

1 Introduction

Demand for unmanned-construction technology to remotely operate heavy machinery at disaster sites appears to be increasing because of the risk of secondary disasters when people directly enter the site to work[1]. However, in unmanned construction, the operator must recognize the complex situation at the site based on limited information from the image captured by onboard camera mounted on the construction machine (cab view) and those images captured by external cameras installed at the work site (external views) [2]. Therefore, the task is very cognitively demanding, and the work efficiency is only approximately 50% compared with that in the boarding operation [3]. To solve this problem, conventional research has been conducted to increase the amount of information by adding other viewpoint cameras and types of sensors mounted on heavy machinery and by providing the operator with multiple displays [4]. However, the increased amount of information requires the operator to select and integrate this information, which can increase the operator cognitive load and reduce work efficiency. In addition, cognitive tunneling may occur under a high cognitive load [5], which makes selection of appropriate information difficult using the conventional multi-display presentation methods. In the present study, we developed a novel visual interface that reduces cognitive tunneling of the operator and supports efficient information selection.

2 Visual Interface Design

2.1 Design Policy

The causes of cognitive tunneling can be mainly divided into two categories.

First, "Visual Momentum," which represents the ease in eye transition among the information presented on multiple screens, is low, which makes information integration difficult. This condition makes the operator suffer from the difficulty of paying attention to other images when the operator attempts to work only with the information obtained from one image type. Second, the operator attention is fixed on the information with high “Visual Saliency” and does not look at other images [6]. In particular, remote operators tend to fix their attention on cab view. Therefore, a visual presentation method that frees the attention of operators from cab view and directs their attention to an appropriate image is needed.

According to the aforementioned conditions, we set two design policies for visual interface: one was to extract and present information that increases visual momentum and the other was to present the information that directs the operator attention. In line with the aforementioned design policies, we proposed the following methods: (1) selecting and presenting external views according to the work state, (2) deriving the optimal viewpoint position for external views, and (3) presenting information that considers the visual field characteristics of humans. We then constructed a visual interface.

2.2 Selection and Presentation of External Views According to the Working State

To increase visual momentum, the image presented to the operator must be related to the current work content. In this study, we developed a method of selecting images that increased the degree of agreement between the landmarks that the operator must recognize as work cues and those contained in the images to define the relationship between the work contents and provided images. The landmarks that must be recognized by the operator were considered to be different depending on the work state. The landmarks that must be recognized in each work state were analyzed based on a teleoperation experiment in a virtual-reality [7] environment and a field survey at Mt. Unzen-Fugen. In particular, the work states were divided into four types: moving, grasping, transporting, and releasing. When moving, it is required to avoid accidental contact with obstacles. So an external camera image that can help the operator understand the distance between the body of heavy machinery and the obstacle is needed. On the other hand, in manipulation work, which includes grasping, transporting, and releasing, an external camera image that shows the positional relationship between the grasped object and the grapple of heavy machinery is necessary. Besides, another external camera image that shows the positional relationship between the grasped object and the placing position is needed.

2.3 Derivation of the Optimal Viewpoint Position for External Views

We experimentally derived the viewpoint position of external views where the operator could most efficiently capture the positional relationship between the work-site environment and heavy machinery and where the work efficiency could be improved [8]. In particular, we compared the working time for grasping and releasing, which required a particularly high-precision work by changing the camera position, and derived the position of the external camera that offered the highest work efficiency. The experimental result demonstrated that the optimum pan angle was $\emptyset = 90°$ and the optimum tilt angle was $\theta = 60°$ in a polar coordinate system using the object to be grasped as a pole. We also believed that the optimal ranges of $\emptyset$ from 60° to 120° and θ from 60° to 90° did not significantly affect the work efficiency.

2.4 Method for Presenting Information Considering the Human Visual Characteristics

To enable the operator to efficiently refer to the information determined by our proposed method, we used a single display that could present external views in the useful field of view and all the images in a stable gazing field of view by considering the human information-reception behavior. In addition, to shift the operator attention, which tended to be fixed on cab view, the screen of external view was made to vibrate at approximately 5 Hz when the external view was selected and presented to encourage the operator to shift his gaze [9]. Figure 1 shows an overview of our developed visual interface.

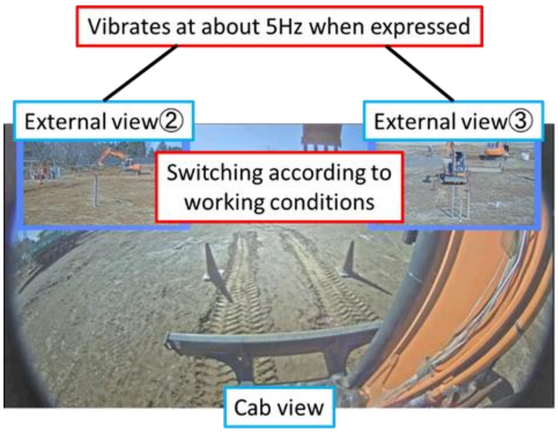

Figure 1 Overview of the visual interface developed

3 Experiment

In order to verify the effectiveness of the proposed visual presentation method, an experiment was conducted using an actual hydraulic excavator (Hitachi ZAXIS 35U, Figure 2) at the outdoor test site of the Public Works Research Institute.

Figure 2 HITACHI ZAXIS 35U

became fully proficient in the operation. The criterion for proficiency was that the error should be within ±5% when the working time was measured and compared with the three nearest working times. The proficiency task was terminated when the abovementioned criterion was met. However, even if the conditions were not satisfied, the task was terminated after a maximum repetition of 10 times. We performed the abovementioned task three times using the conventional method of presenting multiple images on multiple displays (Figure 5) and three times using our novel visual interface. In the new visual interface, the camera switcher decided the work state of the actual device and switched the images. To eliminate the effects of habituation caused by repeating the same task under the same conditions, the order in which the experiments were conducted was divided into two patterns of two people each, as shown in Figure 6.

3.1 Experimental Conditions

The experimental task was set by referring to the model task proposed by the Public Works Research Institute [10]. In this test, the vehicle made one round trip along an L-shaped course, as shown in Figure 4, while grasping and releasing the grasped object (Figure 3) four times. From the starting point shown in Figure 4, the vehicle grasped the object at its front and rotated by approximately 90° to put it down. Then, the vehicle moved along the course, grasped the second grasped object at its front, and rotated by approximately 90° to put it down. Finally, the vehicle moved back to the starting point and repeated the previous steps in the reverse order to complete the task. In addition to cab view, external views from the side are necessary for the operator to obtain depth information in the manipulation work. The positions of the external cameras used in this experiment (External cameras ① to ③) are fixed, and the angle of camera image is changed according to the work state. The angle of camera image is set to be within the range of the optimal pan angle and tilt angle described in section 2.3. Table1 summarizes which cameras were used according to the work state in this experiment.

In this experiment, a familiarization task was conducted before the main test. In the proficiency task, the vehicle started at the corner of the L-shaped course shown in Figure 4, moved along the straight course at the long side, grasped the object at the right (left) side of the excavator from the direction of travel, turned 180°, released the object in place, and finally moved to the starting point.

The subjects were four operators who were skilled in boarding operations. In the experimental flow, the operator first performed the mastery task until he or she

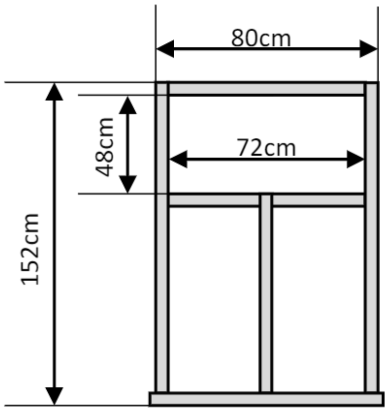

Figure 3 Grasped object

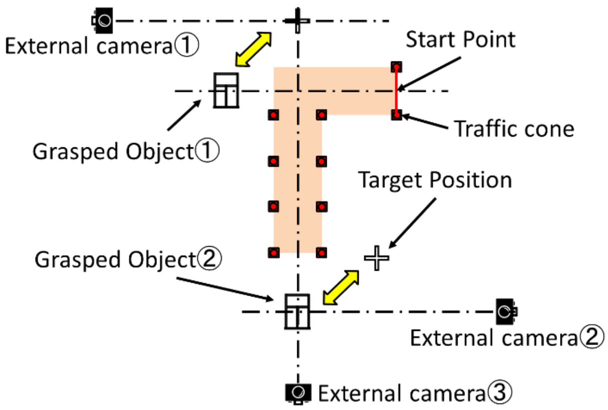

Figure 4 Test field

Figure 5 Multi-Display presentation method

Table1. The camera image presented in each work state

Work states	Camera			
		View that provides depth information	View to confirm proper stopping position	View to avoid false contact
Moving①	Cab view	-	Camera③	Camera①
Grasping①		Camera③	-	-
Releasing①		Camera①	-	-
Moving②		-	Camera②	Camera③
Grasping②		Camera②	-	-
Releasing②		Camera③	-	-

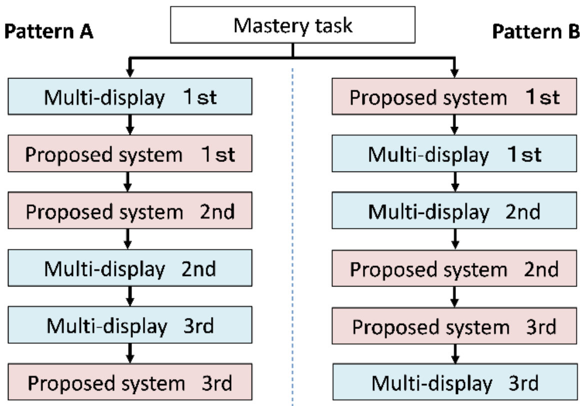

Figure 6 Test flow

We measured the working time, placement error, and number of false contacts as indexes to evaluate the work efficiency. The gaze was evaluated to confirm the effectiveness of the gaze guidance.

In this experiment, placement error was defined as the deviation between the center position of the marker and grasped object. Only the distance between the center of the marker and grasped object was measured, and the rotation angle was not considered. The number of erroneous contacts was defined as the number of contacts between the heavy machinery and the traffic cones that served as the moving course.

3.2 Results and Discussion

3.2.1 Working Time

The results of the Wilcoxon signed-rank sum test showed that no significant difference existed between the conventional and proposed methods in terms of the total working, moving, transporting and releasing times (Figure 7). The results also suggested that the conventional method could better reduce the grasping time than the proposed method.

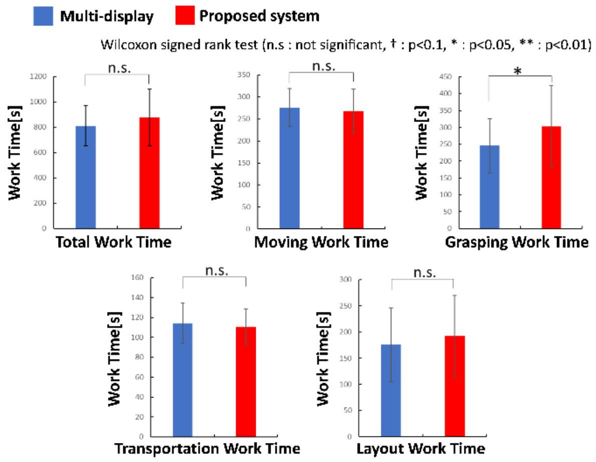

Figure 7 Result on work time

Our study revealed that the working time could not be reduced. In this experiment, the conventional method presented four images to the subject without switching the images, and the angle of view was always the same. Therefore, while working under the condition where all images were constantly visible, we believed that understanding of the images that were not being looked up subconsciously deepened. Thus, no time was required to understand the images when shifting the gaze between images. On the other hand, in the proposed visual interface, because the images were switched according to the work state, time was needed to understand the images every time they were switched. We believed that this “time to understand the image” was a factor that did not lead to reduction in the working time.

3.2.2 Gaze

Figure 8 shows the results of the gazing process. The visual momentum was evaluated according to the ratio of time spent gazing at one image for more than 1.5 s during one task. The cognitive tunneling was evaluated according to the number of times a gazed image was switched from cab view to external views in 1 s to determine if the attention shifted from an image with high visual saliency. The number of times the gazed image was switched from cab view to external views was used to evaluate whether cognitive tunneling was reduced.

The result of the Wilcoxon signed-rank sum test of the four subjects revealed no significant difference between the conventional and proposed methods in terms of the ratio of watching the same image for more than 1.5 s and the number of times of switching from cab view to external views and the number of switching the gazed images.

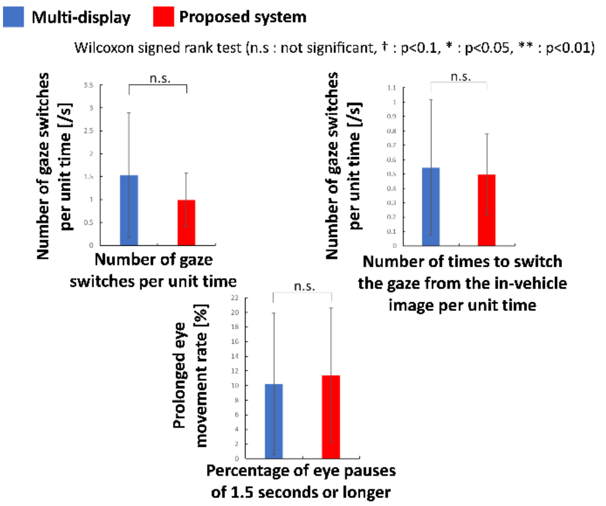

Figure 8 Result on gaze

In the experiments that used the scaled model, we believed that the proposed method significantly reduced the percentage of gaze pauses for 1.5 s or longer and significantly increased the number of gaze changes in each of the previously mentioned evaluation indexes. This point is discussed in this paper. In the experiment that used the scaled model, nine images were presented to the subject using a multi-display presentation method to verify the effect on the work efficiency by reducing the time to select an appropriate image according to the work state. In the experiment using an actual machine, however, the number of presented images was reduced to four to reflect an environment that was closer to that of an actual unmanned-construction site. This process made selection of appropriate images and integration of information among images easier. In addition, whereas the subjects in the experiment that used the scaled model were inexperienced in the operation of an actual machine, the subjects in this experiment were skilled in the operation of boarding machines. Therefore, they were considered skilled in the ability to select the image to be viewed according to the work state. Hence, a possibility existed that no significant difference could be confirmed in each evaluation index.

Next, when we organized the individual differences in terms of the different cognitive characteristics, behavior, and learning mechanisms of the subjects, we were able to identify the operators who were comfortable with the proposed visual interface, as shown in Figure 8. For Subjects B and D, we confirmed that the number of times the gazing image was switched from cab view and that when the gazed image was switched significantly increased using the proposed method. This result suggested that the proposed method can direct the attention of some operators from cab view with high visual saliency to external views with low visual saliency and can reduce cognitive tunneling.

On the other hand, no significant difference was observed in terms of percentage for Subjects B and D in watching the same image for more than 1.5 s in the time on one task (Figure 9). As described earlier, the conventional presentation method always presented four images with the same angle of view. Therefore, the time required to understand the images was likely reduced compared with that in the proposed method. In other words, when the number of images presented was only four, the proposed method could reduce the time required to select an image. However, it required time to understand the image each time it was switched, and the latter had a greater effect on work efficiency. Although the latter process succeeded in directing the gaze to some extent, it did not improve the visual momentum, which could be the reason why the working time was not reduced.

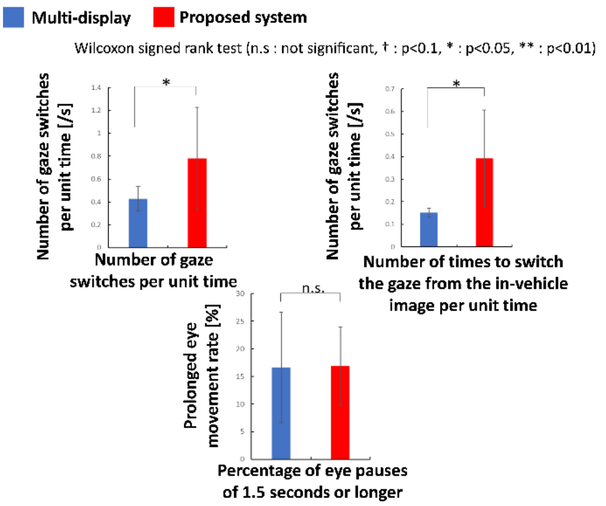

Figure 9 Result on gaze (Subjects B and D only)

3.2.3 Placement Error and Number of False Contacts

Figure 10 shows the results of the placement error and number of erroneous contacts using the Wilcoxon signed-rank sum test.

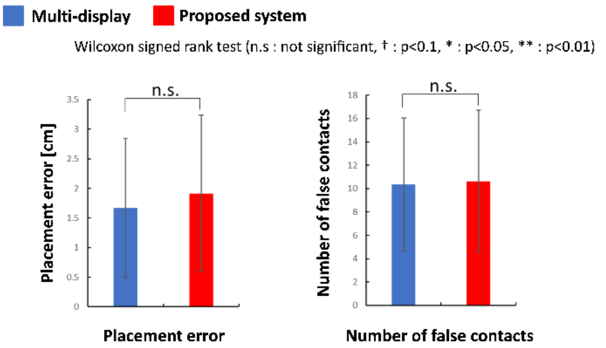

Figure 10 Result on placement error and number of false contacts

Figure 11 Picture angle for each video presentation method

The fact that the placement error and number of false contacts could not be reduced using the proposed method is discussed. Figure 11 shows that the view angles of the images provided by the conventional method and those presented by the proposed method were the same in this experiment. Therefore, when the performance of the placement error and number of false contacts were improved, the proposed method could appropriately assist the selection of the image to be viewed. However, in this experiment, in contrast to the scaled-model experiment, the subjects were likely to already possess a high level of ability to select the images to be viewed. Further, because the number of types of images presented was significantly reduced, selection of the images was easy. In this case, we believe that the effects on the placement error and number of false contacts are less likely to occur.

4 Conclusion

In this study, we developed a visual presentation method that can reduce cognitive tunneling to improve the efficiency of unmanned construction. The cause for cognitive tunneling is that the operator attention is fixed only on the images with low visual momentum and high visual saliency. Therefore, we aimed to improve the visual momentum by including the same landmark in each image. In addition, by presenting external views with low visual saliency in the useful field of view using vibration, the operator attention was directed from cab view with high visual saliency. The results from experiment that used an actual system showed that some operators were able to successfully direct their gaze; however, it did not lead to improvement in the work efficiency. The reason for the lack of improvement in the work efficiency was believed to be the need for time to understand the new external view by switching the image, which is a feature of the proposed interface.

This result indicates that even if the gaze can be directed to the image corresponding to the work state, it cannot solve the essential problem unless it can assist in understanding of the image.

Therefore, in the future, we aim to improve the work efficiency by introducing a system that can assist understanding of spatially discrete images when switching or displaying new images.

5 Acknowledgments

We would like to thank staff members of the Public Works Research Institute for their help with the experiments. We also thank Japan Institute of Country-ology and Engineering for the research grant. This research was supported in part by the Research Institute for Science and Engineering, Waseda University.

References

[1] Ban Y. Unmanned construction system: Present status and challenges. In *International Symposium on Automation and Robotics in Construction*,

pages 241–246, 2002.

[2] Chen J.Y., Haas E.C., Barnes M.J. Human performance issues and user interface design for teleoperated robots. *Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, 37(6):1231–1245, 2007.

[3] Moteki M., Fujino K. and Nishiyama A. Research on proficiency of remote operation of construction machinery. In *Proceedings of the Symposium on Construction Robotics in Japan (SCR),* pages19–24, 2012.

[4] Furuya H., Kuriu N. and Shimizu C. Development of Next-Generation Remote-Controlled Machinery System. In *REPORT OF OBAYASHI CORPORATION TECNICAL RESEARCH INSTITUTE*, 76, 2012.

[5] Dirkin G. Cognitive tunneling: Use of visual information under stress. *Perceptual and Motor Skills*, 56:191–198, 1983.

[6] Thomas L.C. and Wickens C.D. Visual displays and cognitive tunneling: Frames of reference effects on spatial judgments and change detection. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, pages 336–340, 45(4), 2001.

[7] Yang J., Kamezaki M., Iwata H., and Sugano S. Analysis of effective environmental-camera images using virtual environment for advanced unmanned construction. In *International Conference on Advanced Intelligent Mechatronics (AIM)*, pages 664–669, 2014.

[8] Sato R., Kamezaki M., Yamada M., Hashimoto T. and Iwata H. Experimental derivation of the optimum and suitable range in sideways views for digging and releasing in unmanned construction. In *Transactions of the JSME (in Japanese)*, volume 85(876):19-66, 2019.

[9] Shimura M., Suzuki H., Shimomura Y. and Katsuura T. Perception of visualmotion stimulation in peripheral vision during eye fixation. In Japan Society of Physiological Anthropology, 20(2):95-102, 2015.

[10] Moteki M., Yuta S. and Fujino K. The proposal of the model task for efficiency evaluation of the construction work by remote control of a hydraulic excavator. *Journal of JCMA*, 66(8):71-79, 2014.