Experimental Result of Third-person View Generation using Deliberately Delayed Omni-directional Video

A. Sakata[a] , Y. Hada[a] , R. Hojo[a] , M. Munemoto[a] ,
Y. Takeshita[b] , T. Asuma[b] , and S. Kitahara[b]
[a]Kogakuin University
[b]Kumagai Gumi Co., Ltd.
E-mail: am20033@ns.kogakuin.ac.jp, had@cc.kogakuin.ac.jp, am19050@ns.kogakuin.ac.jp, am19055@ns.kogakuin.ac.jp, yoshito.takeshita@ku.kumagaigumi.co.jp, tsubasa.asuma@ku.kumagaigumi.co.jp, skitahar@ku.kumagaigumi.co.jp

Abstract –

This study developed a system that generates video from a view following a robot, using composition CG of the present position of the robot using deliberately delayed video on the robot for displaying the current position and orientation of the robot. The proposed view generation method overcomes the disadvantage of conventional methods because it always has a close-up view, and has the advantage of the video making it easy to understand the area around the robot. In this study, we verified the advantage of this method experimentally using a crawler carrier and the robot, and addressed some problems.

Keywords –

Past image; Remote control; Bird's-eye view

1 Introduction

When a disaster occurs, unmanned construction is used for recovery and investigation using a construction machine in a safe area. In general remote-control and unmanned construction, the pilot controls using the subjective view sent to the pilot's terminal by a camera on a construction machine, or a third-person view sent by a camera installed in the surroundings. Subjective view has the advantage that it can be seen in front of the construction machine at all times, and third-person view has the advantage that it is easy to determine when the construction machine is about to hit obstacles. However, subjective view and bird's-eye view videos have parts that cannot be seen or are difficult to see. The purpose of this research was to develop a system that generates video from a view behind the robot by composition CG of the present position of the robot, using deliberately delayed video from the robot.

2 Related research

2.1 Classification and advantage of camera view

In remote control, camera images are used to understand the environment of the construction machine.

Here, we classify the installation positions of the cameras and the obtained images, and describe the advantages and disadvantages thereof.

- Subjective view

Generally, the camera is installed in the driver's seat of the construction machine, and a subjective image is taken as if the operator is looking forward from the driver's seat. A typical obtained image is shown in Figure 1.

Subjective images are realistic and easy for the operator to imagine, and the device is simple and fully contained in the construction machine; thus, the usage cost is low. However, the general camera can only retrieve limited information because its viewing angle is narrower than the human viewing angle. Thus, there are cases where multiple cameras are installed or a fisheye camera with a wider viewing angle is used [1].

Figure 1. Example of subjective view

- Swingable subjective view

In a previous study, the attachment of a swing mechanism was attached to the camera to allow the operator to check the sides and back to overcome the narrow viewing angle of subjective images [2]. However, it is not possible to change the large view such as leaning forward to check the rear like in a manned vehicle, and the viewing angle is limited. In addition, the swing mechanism is operated separately; VR technology can be used to simplify this[3].

- Fixed third-person view

In unmanned construction, the condition of the construction site and the positional relationship of the construction machine therein are important. Thus, it is common to install cameras on pipe scaffolding or prefabricated scaffolding for a broad third-person view. An image from a typical fixed third-person-view video is shown in Figure 2. The third-person-view video may have a sensational discrepancy between the operation and the movement of the construction machine, because it does not always face the same direction as the construction machine; therefore, it requires skill to operate. For example, the right crawler must be moved to turn the construction machine in Figure 2 to the left. In addition, when the distance between the camera and the construction machine is large, the construction machine in the image may appear small, or another object may appear between the two, hiding the construction machine. Furthermore, the installation cost of the camera is high, and installation of the camera itself may be impossible immediately after a disaster.

Figure 2. Representative fixed third-person view video

- Third-person view from mobile robotsthird-person view

In some disaster response robots, a rod is extended backward on the robot and a camera is installed in a high position on the tip of the rod to obtain a third-person-view video of the robot surroundings. A typical image is shown in Figure 3. In this case, the orientation of the robot and the position and orientation of the camera always match, and it is easy to grasp the surroundings subjectively. Toda et al. obtained a bird's-eye view from the top in real time by mooring a balloon on the robot instead of a stick, and looking down at the robot from the camera installed on the balloon [4]. Nagatani and others also moored a drone on a construction machine [5]. These methods increase the total height, which includes that of the camera. Moreover, they limit the usable environment because it is necessary to physically separate the camera from the center of the robot. Further, there is the drawback that the image is very shaky owing to the effects of rotation, vibration, and wind of the construction machine because the robot moves away from the center of rotation.

The technology of artificially generating a bird's-eye view image by synthesizing a plurality of camera images mounted on a robot on a plane has been applied practically by automobile manufacturers. An example of the one put into practical use is the around view monitor of Nissan Motor Co., Ltd [6]. Sato et al. applied this method to unmanned construction equipment [7]. Shimizu et al. applied this method and made it possible to move the view [8].

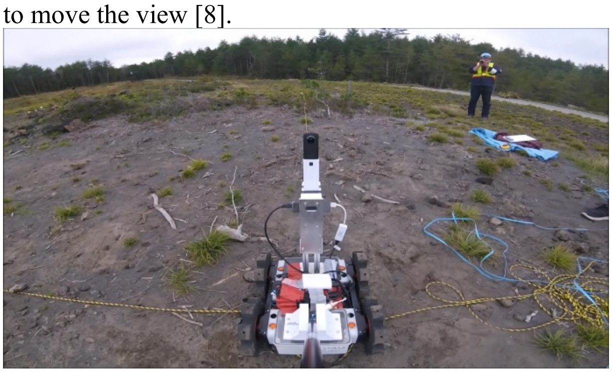


Figure 3. Representative moving third-person view video

2.2 Fixed third-person view using past image method

Shiroma et al. saved a still image acquired by a camera mounted on a robot, and synthesized a CG that matched the position of the robot in the still image in the save history to view the robot from the position where the still image was acquired; they proposed a method to generate an imaginary bird's-eye view [9]. In addition, Kinoshita et al. eliminated the narrow viewing angle by shooting still images using an omni-directional camera [10]. Figure 4 shows the outline and Figure 5 shows the generated third-person-view image.

The fixed third-person-view generation method using previous images involves obtaining the fixed third-person-view view by using previous still images captured by the robot. That is, if the only moving object in the image is the robot, the still image once taken is used as the background, and the moving robot draws the CG according to the position and orientation information. This method has the merit that the amount of communication and the associated communication delay can be reduced because moving image transfer is not required.

In contrast, the image obtained by this method has a fixed third-person-view, and thus there is the problem that the size of the image decreases as the robot moves away. To solve this, it is necessary to retake a still image when a certain distance is reached. In addition, moving objects such as humans and other robots are perceived as stationary, because moving objects other than robots are not assumed.

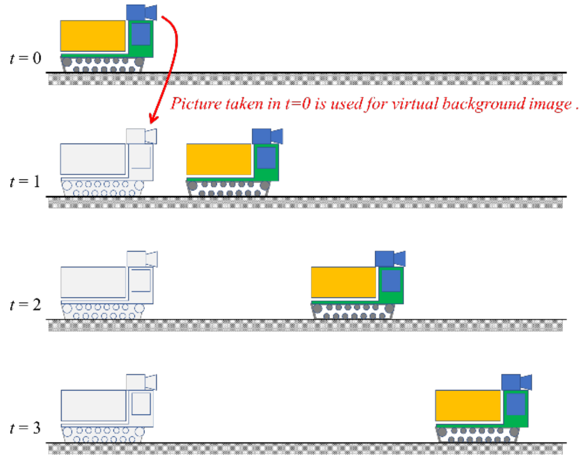


Figure 4. Overview of past image method

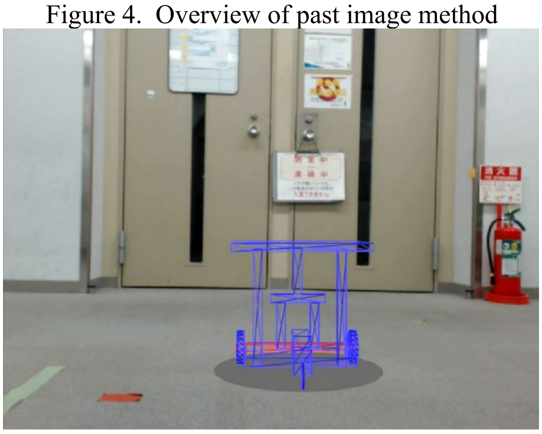


Figure 5. Third-person view image by past image method

3 Tracking third-person view video method

3.1 Proposal of tracking third-person video method

We extended the idea of the fixed third-person-view generation method using past images from an omni-directional camera developed by Kinoshita et al. We used a moving image instead of a still image, and intentionally delaying this to create a CG image at the robot's current position. We proposed a method of generating a moving view in which a camera positioned behind a robot is followed by continuous drawing [11]. Figure 6 shows the outline. In this paper, this is called a tracking bird's-eye view view. In the case of the fixed bird's-eye view image described above, the drawing position of the robot and its size change. However, in this method, it is always drawn in a fixed size in the center of the screen; thus, it has the advantage of being easily drivable. In contrast, in the proposed method, the amount of communication is not reduced because moving images are used. Although the moving object is reflected, it is a delayed moving image and does not correctly reflect the movement of the object.

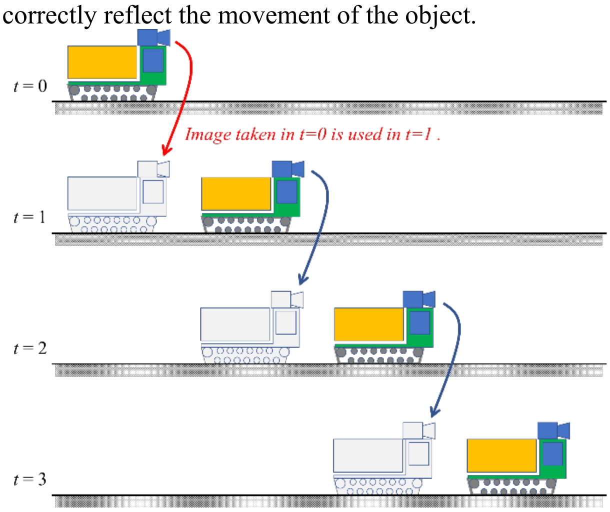


Figure 6. Overview of tracking third-person view method

3.2 Generation method of tracking third-person view video method

The generation procedure of this method is shown as follows.

1. The robot saves the current time and the position and orientation of the robot along with the moving image.
2. A moving image and robot position and orientation that meet the conditions are selected. There are two possible conditions to select.
   A) Fixed time delay—Selection of the latest moving image where the difference between the current time and the shooting time of the image is more than a certain value.
   B) Fixed distance delay—Selection of the latest moving image where the distance between the current position and the shooting position of the image is a certain distance or more.
3. The robot sends the current position/orientation of the robot, the selected moving image, and the position and orientation at the time of image capture to the operator terminal.

4. The CG reflecting the current position and orientation of the robot is superimposed on the moving image received by the operator side, and displayed on the screen. The current direction of the robot is cut from the moving image when drawing so that the CG of the robot is always centered.

3.3 System configuration

Figure 7 shows the system configuration of this research. The upper left side of the dotted line shows the processing performed by the robot side, and the upper right side shows the processing performed by the operator side. These are the contents of the software. The lower part of the dotted line shows the required hardware. The system on the operator side calculates the relative position and orientation from the shooting point from the current position and orientation received from the system on the robot side and the position and orientation at the time of shooting the moving image, and synthesizes the CG of the robot with the moving image. The positional relationship between the self-position estimation sensor, the omni-directional camera, and the robot must be obtained before the experiment. The position where CG is combined with the moving image is calculated as follows. Figure 8 shows an example of robot behavior. Position/posture P_{pc} of the camera at the time of shooting when the position/orientation of the robot center at the time of shooting is the origin, position/posture P_{pr} of the robot center at the time of shooting, sensor position/posture P_{ps} at the time of shooting, current position/posture P_{nr} of the robot center, The current position/orientation P_{ns} of the sensor is defined as follows.

$$P_{pc} = (x_{pc}, y_{pc}, z_{pc}, yaw_p, pitch_p, roll_p)$$
$$P_{pr} = (x_{pr}, y_{pr}, z_{pr}, yaw_p, pitch_p, roll_p)$$
$$P_{ps} = (x_{ps}, y_{ps}, z_{ps}, yaw_p, pitch_p, roll_p)$$
$$P_{nr} = (x_{nr}, y_{nr}, z_{nr}, yaw_n, pitch_n, roll_n)$$
$$P_{ns} = (x_{ns}, y_{ns}, z_{ns}, yaw_n, pitch_n, roll_n)$$

The relative position of the position and orientation of the camera, the position and orientation of the robot center, and the position and orientation of the sensor are fixed because they are fixed to the robot. Also, assuming that the directions of the camera, sensor, and robot are the same, the yaw, pitch, and roll values are the same. Therefore, the position and orientation P_{rs} of the sensor viewed from the center of the robot and the position and orientation P_{sc} of the camera viewed from the sensor are defined as follows.

$$P_{rs} = (x_{rs}, y_{rs}, z_{rs}, 0,0,0)$$
$$P_{sc} = (x_{sc}, y_{sc}, z_{sc}, 0,0,0)$$

When the robot moves as shown in Fig. 8, the vector from the camera at the time of shooting to the center of the robot after the movement is obtained by the following procedure.

1. Obtain the position and orientation of the camera during shooting from the position and orientation of the sensor during shooting

$$P_{pc} = P_{ps} + P_{sc}$$

2. Obtain the coordinates of the current robot center from the current sensor position and orientation and the sensor position and orientation seen from the robot center

$$P_{nr} = P_{ns} - R_y R_p R_r P_{rs}$$

However, $R_y R_p R_r$ is the following formula.

$$R_y = \begin{bmatrix} \cos\left(yaw_n - yaw_p\right) & -\sin\left(yaw_n - yaw_p\right) & 0 & 0 & 0 & 0 \\ \sin\left(yaw_n - yaw_p\right) & \cos\left(yaw_n - yaw_p\right) & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$R_p = \begin{bmatrix} \cos\left(pitch_n - pitch_p\right) & 0 & \sin\left(pitch_n - pitch_p\right) & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ -\sin\left(pitch_n - pitch_p\right) & 0 & \cos\left(pitch_n - pitch_p\right) & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$R_r = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \cos\left(roll_n - roll_p\right) & -\sin\left(roll_n - roll_p\right) & 0 & 0 & 0 \\ 0 & \sin\left(roll_n - roll_p\right) & \cos\left(roll_n - roll_p\right) & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

3. Obtain the current position/orientation P_{pcns} of the robot center with the position/orientation of the camera as the origin

$$P_{pcns} = P_{ns} - P_{pc}$$

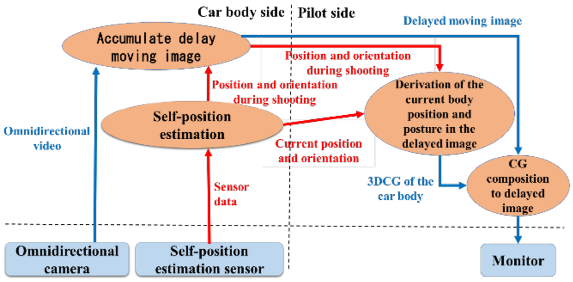
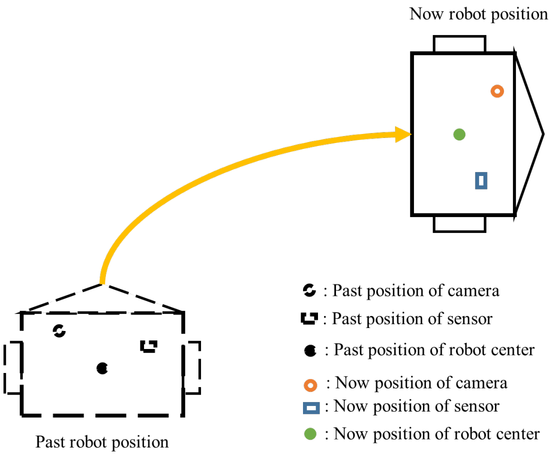


Figure 7. System Configuration



4 Figure 8. An example of robot behavior Experiment of generating a third-person view video

4.1 Video generation experiment using construction machinery

4.1.1 Purpose

Using the proposed method, we attempted to generate a tracking third-person-view image from the moving image of the omni-directional camera acquired while traveled the crawler carrier equipped with the implemented robot side system and the information of RTK positioning.

4.1.2 Data acquisition robot side system

Figure 9 shows the robot system used to acquire the data. The installed camera is the omni-directional camera which name is " RICOH THETA V ".

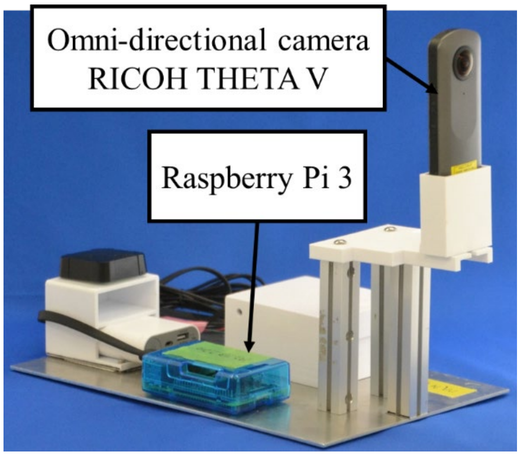


Figure 9. Data acquisition robot side system

4.1.3 Content

The robot side system shown in Figure 10 was installed in a crawler carrier IC 120 manufactured by KATO WORKS. Figure 10 shows the dimensions of the crawler carrier, and Figure 11 shows the camera installation position. The position of the crawler carrier was determined using the data of RTK positioning.position In this experiment, only the position of the crawler carrier and the moving image of the omni-directional camera were acquired during operation, and the video was generated offline. The condition for moving image selection during image generation was the fixed time delay described in 2(a) in section 3.2.

Figure 10. IC120 dimensions

Figure 11. Camera's position

4.1.4 Results and discussion

The video in Figure 12 was generated offline from the data acquired during the experiment. The video is available at https://youtu.be/HMjuIucaNg4. The generated image confirms that the rough movement of the CG of the construction machine matches the movement of the actual construction machine, and that the view follows the CG. As a result, we demonstrated that it is possible to generate a third-person view image from the image of the vehicle-mounted camera. There was a problem in that the view sometimes caught up with CG when the construction machine speed decreased. This is probably because the delay time was constant and the distance to the camera changed depending on the speed of the crawler carrier. There was a further problem that the CG skidded and oscillated. It is considered that this is because the shooting time of the moving image and the acquisition time of the position and orientation were not correctly synchronized.

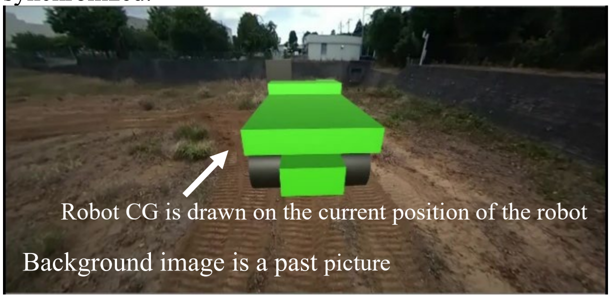


Figure 12. Video generated in an experiment using a crawler carrier

4.2 Video generation experiment using a small dolly

4.2.1 Purpose

Based on the experiment detailed in the previous section, we attempted to solve the problem by improving the system. The video was generated in real time to check for practical problems.

4.2.2 Implementation on a robot for running experiments

We implemented the system on Beego, a small indoor vehicle shown in Figure 13. The installed camera is the omni-directional camera which name is " RICOH THETA S ".The odometry obtained from the number of tire rotations was used to acquire the position and orientation of Beego.

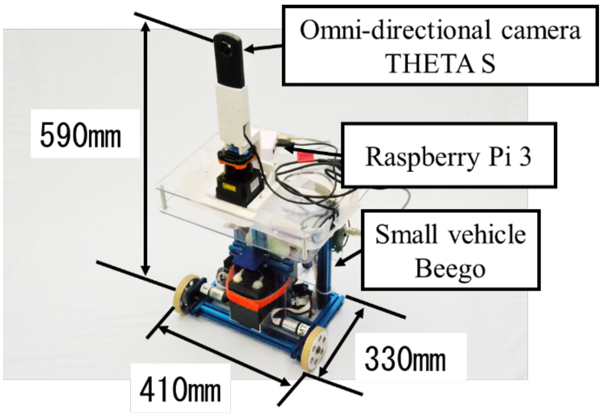


Figure 13. Mounted small vehicle

4.2.3 Content

We generated a tracking third-person-view image with the improved system. In the experiment, the vehicle traveled from the position shown in Figure 14 along a 2-m square perimeter, drawn as a dashed black line. The turns were made on the spot or without stopping. The image selection condition for moving image generation was the constant distance delay (75 cm) described in Procedure 2(a) in Section 3.2.

Figure 14. Indoor experimental environment

4.2.4 Results and discussion

An image from the video generated in real time while driving is shown in Figure 15; the video is available at https://youtu.be/GBiFHKc6bh0. From the generated image shows that the camera position changed according to the robot's moving and stopping, we confirm that the position change of the camera also changed with the speed and stopping of the robot. From this, we succeeded in keeping the distance between the robot and the camera constant. However, the CG was drawn at a different position to the actual position. When the CG bent inward from the actual robot position, especially when it turned smoothly without stopping, it was inflated and drawn outside. A visual check confirmed that the actual robot position and the CG position were shifted by a maximum of approximately 106 cm. And, the stop position was offset by about 5 cm in the direction of travel. It is considered that this is because the operation of the program is slow, and the delay of the program is unintentionally added to the intentional delay.

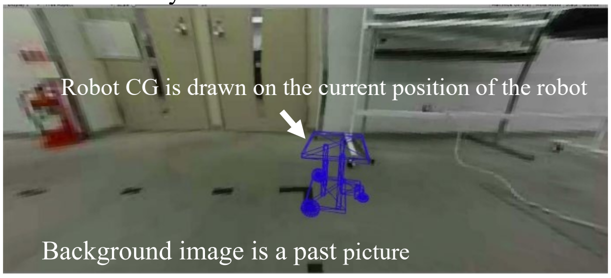


Figure 15. Video generated in an experiment using a small vehicle

5 Summary and outlook

This paper described a method to generate a video that follows the robot from the back, using the camera on the robot, by intentionally superimposing the CG of the robot on the delayed video. We implemented a tracking view generation method. We conducted experiments using two types of robots—a construction machine and a small vehicle—and confirmed the basic operation. From the experimental results, we prove that the third-person view image can be generated from the image of the camera on the robot according to the theory, and it was confirmed that the distance between the robot and the camera can be fixed. A mounting problem caused a phenomenon in which the actual position of the robot and the position of the CG caused an error and the CG skid. As a result, in addition to the intentionally generated delay, there was an unintentional delay in the program.

In the future, we plan to reduce the delay in the system to more accurately superimpose CG. In addition, we aim to demonstrate the superiority of this research by performing a subject experiment using the generated tracking bird's-eye-view image and other images.

References

[1] Kyohei Yoshida, Fuminori Hibino, Yasutake Takahashi, and Yoichiro Maeda, “Development and Evaluation of Human Pointing Navigation System Using Spherical Vision System on Mobile Robot”, Proceedings of 27th Fuzzy System Symposium, WC3-1, 2011. [in Japanese]

[2] Tomoki Tajiri, Rei Ogami, Yogo Takada, and Tadao Kawai, “Supporting of Safe Walking at a Crossing by Using a Small Mobile Robot with Cameras”, Proceedings of the 2012JSME Conference on Robotics and Mechatronics, 1A2-G09, 2012. [in Japanese]

[3] Takashi Kato. and Jun Ueno. “Development of realistic type remote control system using the HMD”. Proceedings of The 16th Symposium on Construction Robotics in Japan, p2-4, 2016. [in Japanese]

[4] Koki Toda, Naoya Maeda, Motoko Kanegae, Yuta Sugiura, Masaahiko Inami, and Maki Sugimoto, “Real time visual operational support interface to display a third-person view of a camera suspended from a floating ballon”, Proceedings of JSME annual Conference on Robotics and Mechatronics 2013, 1P1–P01, 2013. [in Japanese]

[5] Chikushi S, Moriyama Y, Fujii H, Tamura Y, Yamakawa H, Nagatani K, Sakai Y, Chiba T, Yamamoto S, Chayama K, Yamashita A and Asama H (2020), "Automated Image Presentation for Backhoe Embankment Construction in Unmanned Construction Site", In IEEE/SICE International Symposium on System Integration (SII)., pp. 22-27, 2020.

[6] “Intelligent Around View Monitor | NISSAN | Technology Development initiatives”. https://www.nissan-global.com/JP/TECHNOLOGY/OVERVIEW/iavm.html .

[7] Takaaki Sato, Alessandro Moro, Hiromitsu Fuji, Kazuya Sugimoto, Akira Nozue, Yoichi Mimura, Katsumi Obata, Atsushi Yamashita, and Hajime Asama, “Development of Virtual Bird’s-eye View System in Unmanned Construction,” Proceedings of the 19th Robotics Symposia, 346-352, 2014. [in Japanese]

[8] Seiya Shimizu. Masami Mizutani. and Tohru Tsuruta. “Development of Wraparound View System for Vehicles”. Proceedings of The Journal of The Institute of Image Information and Television Engineers Volume 68 Issue 1, pp. J24-J29, 2014. [in Japanese]

[9] Naoji Shiroma, Georges Kagotani, Maki Sugimoto, Masahiko Inami, and Fumitoshi Matsuno, “A Novel Teleoperation Method for a Mobile Robot Using Real Image Data Records,” Proceedings of the 2004 IEEE International Conference on Robotics and Biomimetics, pp. 233-238, 2004.

[10] Himuro Kinoshita, and Yasushi Hada， “Research on Remote Control of Robots Using Omnidirectional Past Images,” The 34th Annual Conference of The Robotics Society of Japan, 2D1-05, 2016. [in Japanese]

[11] Yuya Kaneko and Yasushi Hada, “Third person view generation for remote control robot,” Japan Society of Mechanical Engineers Kanto Student Council 58th Student Graduation Research Presentation, 206, 2019.