Paper: Rb*--**-****:**

Arbitrary Viewpoint Visualization for Teleoperated Hydraulic Excavators

Tatsuki Nagano*, Ryosuke Yajima*, Shunsuke Hamasaki*, Keiji Nagatani*, Alessandro Moro*, Hiroyuki Okamoto, Genki Yamauchi***, Takeshi Hashimoto***, Atsushi Yamashita*, and Hajime Asama***

*The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
E-mail: nagano@robot.t.u-tokyo.ac.jp
**RITECS Inc.
403 3-5-11 Shibasaki-cho, Tachikawa-shi, Tokyo 190-0023, Japan
***Public Works Research Institute
1-6 Minamihara, Tsukuba-shi, Ibaraki 300-2621, Japan
[Received 00/00/00; accepted 00/00/00]

In this paper, we propose a visualization system for the teleoperation of excavation works using a hydraulic excavator. An arbitrary viewpoint visualization system is a visualization system that enables teleoperators to observe the environment around a machine by combining multiple camera images. However, when applied to machines with arms (such as hydraulic excavators), a part of the field of view is shielded by the image of the excavator's arm; hence, an occlusion occurs behind the arm. Furthermore, it is difficult for teleoperators to understand the three-dimensional (3D) condition of the excavating point because the current system approximates the surrounding environment with a predetermined shape. To solve these problems, we propose two methods: (1) a method to reduce the occluded region and expand the field of view, and (2) a method to measure and integrate the 3D information of the excavating point to the image. In addition, we conduct experiments using a real hydraulic excavator, and we demonstrate that an image with sufficient accuracy can be presented in real-time.

Keywords: arbitrary viewpoint image, visualization, fish-eye camera, hydraulic excavator, teleoperation

1. Introduction

It is essential to recover expeditiously from frequently occurring natural disasters, such as earthquakes and volcanic eruptions. However, the environment close to disaster sites is often dangerous, and it is important to ensure the safety of workers on disaster sites. Accordingly, it is expected that recovery work is performed by the teleoperation of mobile robots and construction machines from safe areas[1, 2]. Generally, during teleoperation, the operator operates machines while watching a real-time image of real environments taken by a camera. However, the real-time image taken by a simple camera has a narrow field of view and an occlusion. Consequently, the work efficiency of the teleoperation process is decreased compared to onboard operation. Therefore, it is important to improve the image presented to the teleoperator to enhance operability[3].

In this study, we focused on excavation work using a hydraulic excavator. During the excavation work, the image of the area around the machine body must be presented for the teleoperator to confirm that safety protocols are satisfied during movement or turning. Furthermore, the image of the excavating point near the bucket also has to be presented. One of the methods employed to solve such problems is the arbitrary viewpoint visualization system [4–6], which allows teleoperators to watch a robot or construction machine from a freely configured third-person perspective. The system is composed of multiple fish-eye cameras that are installed to capture the surrounding environment, as shown in **Fig. 1**. The fish-eye cameras have a wide-angle field of view, and can take images in all directions with a small number of cameras. By projecting images taken by multiple fish-eye cameras onto a dome-shaped model that assumes the surrounding environment, it is possible to view a 360° omnidirectional image of the environment on a single screen, as shown in **Fig. 2**.

However, there are two problems when the arbitrary viewpoint visualization system is applied to the excavation work with a hydraulic excavator. First, the surrounding environment is shielded by the excavator's arm, meaning the rear section of the arm becomes an occlusion. **Fig. 3** clearly shows an example of this problem that occurred when the system was applied to a hydraulic excavator, and an arbitrary viewpoint image was generated. This problem occurs because this arbitrary viewpoint visualization system is based on the assumption that the fish-eye cameras always capture only the surrounding en-

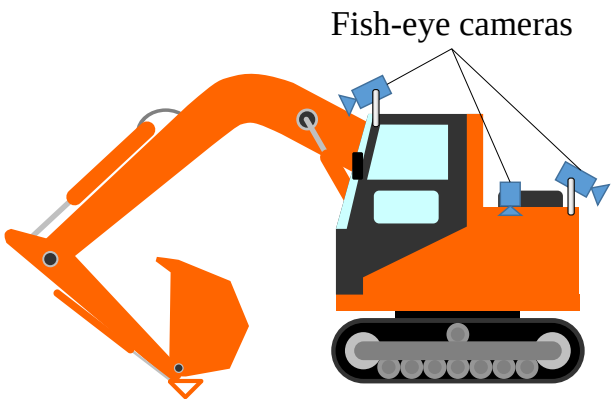

Fig. 1. : Arbitrary viewpoint visualization system.

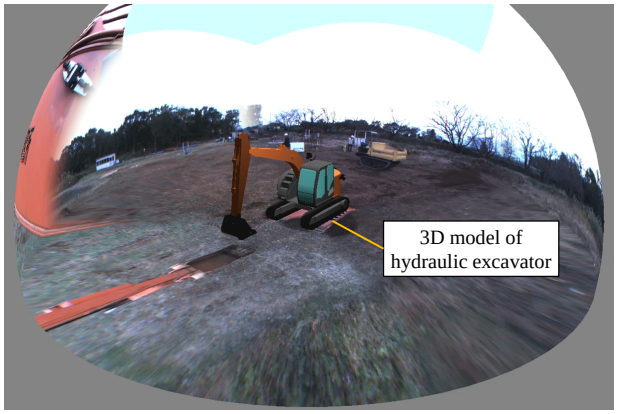

Fig. 2. : Example of arbitrary viewpoint image.

vironment, even though the excavator's arm is in the field of view of the fish-eye cameras. At disaster sites, there may be obstacles behind the occlusion caused by the arm, such as dump trucks, other excavators, or robots. Accordingly, if the current system is applied to a hydraulic excavator, there is a risk of collision with obstacles. The second problem is that the three-dimensional (3D) information of the excavating point cannot be presented to the teleoperator. Therefore, it becomes difficult for the teleoperator to ascertain the 3D condition of the excavating point (such as information about the deposition of sediments and the depth of the hole), which are necessary to determine the position of the bucket during the excavation. The reason for this problem is that the current arbitrary viewpoint visualization system generates images that are based on the assumption that the surrounding environment can be approximated to a predetermined dome-shaped model. In the conventional method, the excavating point is approximated to a flat plane.

During excavation work by teleoperation of a hydraulic excavator, it is effective to use an arbitrary viewpoint visualization system for the teleoperator. However, to ensure safe and efficient teleoperation, the previously mentioned two problems must be solved. Therefore, the purpose of this study is to solve these two problems and to construct an extended arbitrary viewpoint visualization system for

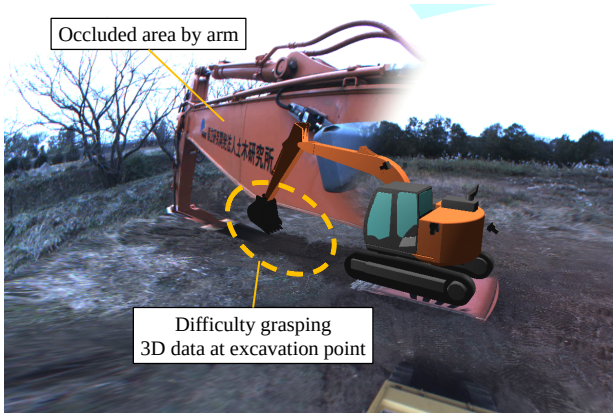

Fig. 3. : Two problems of arbitrary viewpoint images for excavators generated using conventional method[4–6].

the teleoperation of excavation work with a hydraulic excavator. Specifically, to solve the problem of the occlusion by the arm, we propose a method to generate images only from fish-eye camera images that do not capture the excavator's arm. Moreover, to solve the problem of the lack of 3D information of the excavating point, the 3D information measured by an RGB-D sensor is synthesized with an arbitrary viewpoint image.

This paper is organized as follows. In Chapter 2, we describe the overview of the proposed system. Then, in Chapter 3, we describe the method of generating an arbitrary viewpoint image in which the occlusion caused by the excavator's arm is removed, and the verification experiments are also described. In Chapter 4, we describe the system for presenting 3D information of the excavating point using an RGB-D sensor, and the verification experiments are also described. The conclusions and future works are given in Chapter 5.

2. Overview of Proposed Method

2.1. Approach to Solve the Two Problems

First, to solve the problem of the occlusion by the arm, an arbitrary viewpoint image is generated only from fish-eye camera images that do not capture the excavator's arm. To determine whether an arm is included in the image, the positional relationship between the arm and the camera is calculated from the joint angle information and the shape model of the machine body. Using this method, it is possible to expand the field of view of the generated arbitrary viewpoint image by removing the occlusion caused by the excavator's arm.

Second, to solve the problem of the lack of 3D information of the excavating point, we used an RGB-D sensor attached to the arm. The RGB-D sensor can measure the depth information of each pixel in addition to the normal RGB image. The information measured by this sensor is synthesized with an arbitrary viewpoint image and presented to the teleoperator.

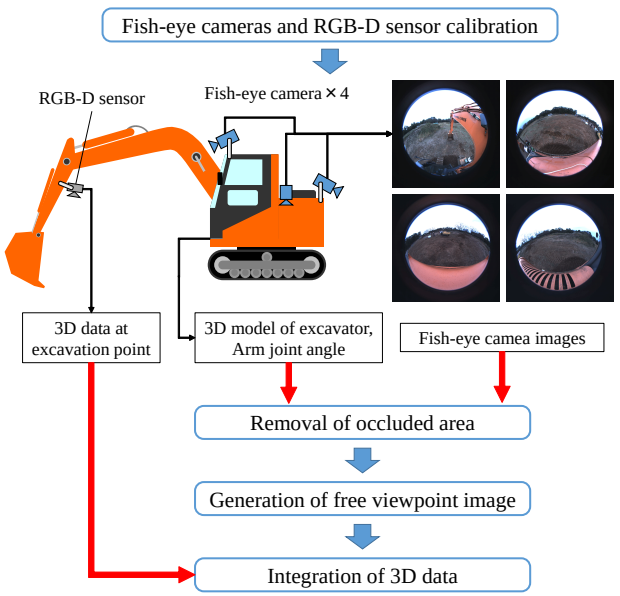

Fig. 4. : Image generation process of proposed method.

2.2. Proposed System and Image Generation Process

This study proposes the extended arbitrary viewpoint visualization system, where the area occluded by an arm can be reduced, and 3D information about the excavating point can be obtained. The proposed system consists of four fish-eye cameras placed on the excavator body in different directions, and an RGB-D sensor attached to the arm. Based on the proposed system, the image generation process is as shown in **Fig. 4**. First, the reflection of the arm in an image is determined from the position and attitude of the cameras estimated by calibration in advance, a 3D shape model of a hydraulic excavator, and the joint angle of the arm. As a precondition, note that the joint angle of the arm can be obtained from sensors equipped on the hydraulic excavator. Second, from the results of the assessment, a dome-shaped arbitrary viewpoint image is generated using only images in which the arm is not reflected. Finally, the ground shape information obtained by the RGB-D sensor is integrated into the arbitrary viewpoint image.

2.3. Coordinate definition

As preparation, the three coordinate systems used in this paper are defined: the base coordinate system Σ_{B} of an excavator, a fish-eye camera coordinate system $\Sigma_{\mathrm{F_i}}(i = 1,2,3,4)$, and an RGB-D sensor coordinate system Σ_{D}. Here, a coordinate in a 3D coordinate system Σ_{N} for a point P within a space is expressed as ${}^{\mathrm{N}}\boldsymbol{p} = [x_{\mathrm{N}}, y_{\mathrm{N}}, z_{\mathrm{N}}]^{\top}$. Moreover, it is also expressed as ${}^{\mathrm{N}}\tilde{\boldsymbol{p}} = [x_{\mathrm{N}}, y_{\mathrm{N}}, z_{\mathrm{N}}, 1]^{\top}$ using a homogeneous coordinate. Then, the transformation matrix $\mathbf{H}$ from the 3D coordinate system Σ_{N} to $\Sigma_{\mathrm{N'}}$ can be expressed as the following equation (1) using a rotation matrix $\mathbf{R}$ of 3 rows × 3 columns, and a 3D translation vector $\mathbf{t}$.

$$ {}^{\mathrm{N'}}\tilde{\boldsymbol{p}} = \mathbf{H}\,{}^{\mathrm{N}}\tilde{\boldsymbol{p}}\,, \quad (1) $$

$$ \mathbf{H} = \left[\begin{array}{c|c} \mathbf{R} & \mathbf{t} \\ \hline \mathbf{0} & 1 \end{array}\right] \quad (2) $$

When fish-eye cameras and an RGB-D sensor are used, it is necessary to estimate the transformation matrix between the three coordinate systems by performing calibration in advance.

3. Generation of Arbitrary Viewpoint Image With Arm Removed

To solve the first issue, an arbitrary viewpoint image is generated in which the arm of an excavator is removed. In this chapter, the generation method and experiments using a real hydraulic excavator in which the proposed system is implemented are presented, in addition to an evaluation of the generated images.

3.1. Method

The proposed method can be divided into three elements: fish-eye camera calibration, image generation, and the removal of the occluded area.

3.1.1. Fish-eye Camera Calibration

As the four fish-eye cameras that constitute the arbitrary viewpoint visualization system are retrofitted cameras, their positions and attitudes relative to the excavator are not known. Therefore, they first need to be calibrated. For a fish-eye camera, calibration involves deriving the transformation matrixes from the base coordinate system Σ_{B} of an excavator to each fish-eye camera coordinate system $\Sigma_{\mathrm{F_i}}$ $(i = 1,2,3,4)$.

First, the relationship between the fish-eye camera coordinate system Σ_{F} and the fish-eye camera image coordinate system Σ_{f} is considered. In this study, a fish-eye camera having an angle of view of approximately 180° is used to secure the wide view field. Fish-eye cameras follow a model that differs from the perspective projection of general pinhole cameras, and the images taken by the fish-eye cameras have characteristic distortion. Therefore, the correspondence between a point ${}^{\mathrm{F}}\boldsymbol{p} = [x_{\mathrm{F}}, y_{\mathrm{F}}, z_{\mathrm{F}}]^{\top}$ in the fish-eye camera coordinate system Σ_{F} and a point ${}^{\mathrm{f}}\boldsymbol{m} = [u_{\mathrm{f}}, v_{\mathrm{f}}]^{\top}$ in the fish-eye camera image coordinate system is derived from the following equation (3) using the method proposed by Scaramuzza et al.[7, 8]

$$ {}^{\mathrm{F}}\boldsymbol{p} = \begin{bmatrix} x_{\mathrm{F}} \\ y_{\mathrm{F}} \\ z_{\mathrm{F}} \end{bmatrix} = a \begin{bmatrix} u_{\mathrm{f}} \\ v_{\mathrm{f}} \\ f(\rho) \end{bmatrix} \left(\rho = \sqrt{u_{\mathrm{f}}^2 + v_{\mathrm{f}}^2} \right). \quad (3) $$

a is a scalar coefficient. $f(\rho)$ is a function that represents the effect of lens distortion, and it depends only on the distance from the center of an image. Using the correspondence of this equation, the characteristic distortion of fish-eye cameras can be corrected, and the image can be

translated to an image such as the perspective projection image of general pinhole cameras.

Next, based on the study by Sato et al., the fish-eye cameras are calibrated using a square calibration pattern the size of which is known. By taking an image of the same pattern placed on the ground from multiple cameras and detecting the four vertexes of the square, the transformation matrix from the world coordinate system Σ_W to the fish-eye camera image coordinate system Σ_f can be derived. Moreover, the transformation matrix from the world coordinate system Σ_W to each fish-eye camera coordinate system Σ_{F_i} $(i = 1,2,3,4)$ can be derived using the above correspondence between the fish-eye camera coordinate Σ_F and the fish-eye camera image coordinate Σ_f. Furthermore, the world coordinate system Σ_W and the base coordinate system Σ_B are aligned manually so that the z-axis and the pivot of the excavator, and the x-y plane and the ground surface are fitted. From the above, the transformation matrix from the base coordinate system Σ_B of the excavator to each fish-eye camera coordinate system Σ_{F_i} $(i = 1,2,3,4)$ can be derived.

3.1.2. Generation of arbitrary viewpoint image

An arbitrary viewpoint image is generated by projecting fish-eye camera images on a dome-shaped model that is assumed to be the surrounding environment. The dome-shaped model consists of a set of numerous triangular meshes that are divided into small pieces, and it is represented by a hemisphere that has as its center the origin of the base coordinate system that shows the distant place and a flat plane that shows the ground near the excavator.

An image that is projected at a point P on the dome-shaped model is determined by the position ${}^{f}\boldsymbol{p}$ of the point P on the fish-eye camera image coordinate system Σ_f, which is calculated from the calibration result. Here, there are cases where images from multiple fish-eye cameras can be projected to the point P because fish-eye cameras have a wide field of view and they are overlapped. In these cases, the use of a fish-eye camera image is determined according to the distance on the image from the center of the image to the point ${}^{f}\boldsymbol{p}$. This is because fish-eye cameras have a property whereby the center part of the image has a higher spatial resolution and a smaller distortion than the peripheral part. In the model of the fish-eye camera that is used in this study, the distance on the image from the center of the image to the point ${}^{f}\boldsymbol{p}$ is proportional to the angle of incidence to the fish-eye camera. The angle of incidence to each camera can be calculated by performing the following equation (4) using coordinates on the fish-eye camera coordinate systems Σ_{F_i}.

$$\theta_i = \arctan \frac{\sqrt{x_{F_i}^2 + y_{F_i}^2}}{z_{F_i}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

In this study, to determine the image to be used, a weighting based on the angle of incidence θ_i is designed according to the method proposed by Komatsu [14]. An image that is projected at the point P is determined by the alpha

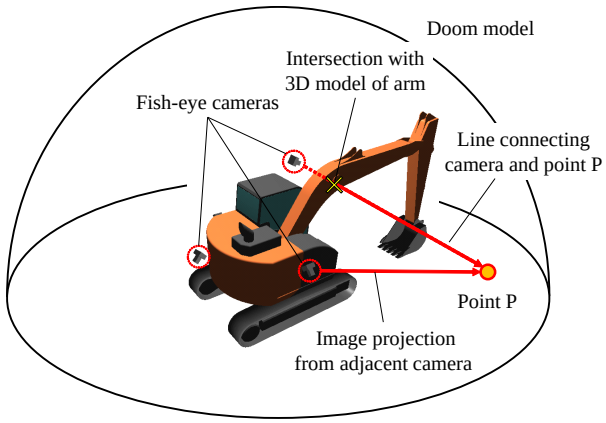

Fig. 5. : Process of removing occluded areas.

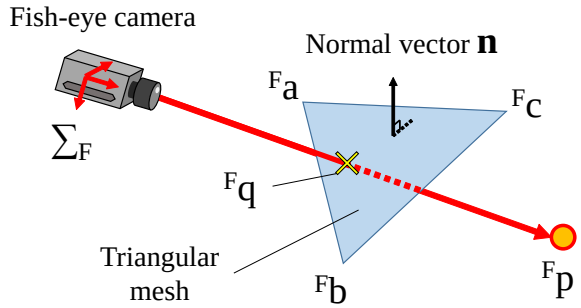

Fig. 6. : Intersection judgement.

blending of multiple fish-eye camera images using the designed weighting. Specifically, the image at the point P is generated from the equations (5) and (6) by using the images I_k and I_l from the fish-eye cameras having the smallest and the second smallest angles of incidence θ_i.

$$I = \alpha I_k + (1-\alpha) I_l \, , \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

$$\alpha = \begin{cases} 1 & (\theta_l - \theta_k \geq \theta_{th}) \\ (1 + \sin \frac{\theta_l - \theta_k}{2\theta_{th}} \pi)/2 & (otherwise) \end{cases} \quad . \quad . \quad (6)$$

θ_{th} represents a parameter that is used to determine the range of the blending, and it means that only the image from the fish-eye camera having the smallest angle of incidence is used when the difference between the angle of incidence is more than θ_{th}.

3.1.3. Removal of occluded areas

The reason why the arm is reflected in the images is that it has been assumed that the fish-eye cameras take images of only the surrounding environment during the above-mentioned image generation method. In this study, the reflection of an arm at a point P on the dome-shaped model is determined based on a 3D shape model of an excavator and each joint angle data of the arm. As shown in **Fig. 5**, the arm is reflected in the image that is projected to the point P when the ray connecting the point P and the origin of the fish-eye camera coordinate system

$\Sigma_{F_i}(i = 1,2,3,4)$ crosses the shape model of the excavator. Therefore, the reflection of the arm can be removed by choosing only images from fish-eye cameras where the ray does not cross the shape model. The invisible region behind the arm in the conventional method becomes visible by using this method, and the arbitrary viewpoint image in which the visible region is expanded can be generated. However, there are also some regions where the images without the reflection cannot be obtained from any fish-eye cameras depending on their distribution. In this study, such regions are painted black.

The details of the method used to determine the intersection are as follows. With respect to the study pertaining the removal of unnecessary objects from an image has been studied [15], the method to prevent forgetting to remove and to ease unnatural image composition is to set the removal region to be larger than the actual size, instead of calculating it from the contour accurately. Therefore, this study considers a polyhedron that consists of multiple triangular meshes and that includes the shape model of the arm. When the intersection judgment is applied to all triangular meshes, and one of these is determined to be crossing, it is determined that the ray crosses the model.

Here, as shown in **Fig. 6**, in a fish-eye camera coordinate system Σ_F, an assessment is made to determine whether a ray connecting a point ${}^F\boldsymbol{p}$ and the origin crosses a triangular mesh that consists of three points ${}^F\boldsymbol{a}, {}^F\boldsymbol{b}, {}^F\boldsymbol{c}$. When the perpendicular of the triangular mesh is represented by $\boldsymbol{n}$, the point of intersection ${}^F\boldsymbol{q}$ between the ray and the flat plane including the triangular mesh is represented by the following equation (7).

$$ {}^F\boldsymbol{q} = \frac{(-{}^F\boldsymbol{a}) \cdot \boldsymbol{n}}{(-{}^F\boldsymbol{a}) \cdot \boldsymbol{n} + ({}^F\boldsymbol{p} - {}^F\boldsymbol{a}) \cdot \boldsymbol{n}} \, {}^F\boldsymbol{p} \quad \ldots\ldots \quad (7) $$

When there exists an intersection point ${}^F\boldsymbol{q}$ inside the triangular mesh, i.e., when positive or negative of three equations (8), (9) and (10) match, it can be determined that the ray crosses the triangular mesh.

$$ (({}^F\boldsymbol{b} - {}^F\boldsymbol{a}) \times ({}^F\boldsymbol{q} - {}^F\boldsymbol{a})) \cdot \boldsymbol{n}, \quad \ldots\ldots \quad (8) $$

$$ (({}^F\boldsymbol{c} - {}^F\boldsymbol{b}) \times ({}^F\boldsymbol{q} - {}^F\boldsymbol{b})) \cdot \boldsymbol{n}, \quad \ldots\ldots \quad (9) $$

$$ (({}^F\boldsymbol{a} - {}^F\boldsymbol{c}) \times ({}^F\boldsymbol{q} - {}^F\boldsymbol{c})) \cdot \boldsymbol{n} \quad \ldots\ldots \quad (10) $$

3.2. Experiments

To verify whether the proposed system can generate arbitrary viewpoint images in real-time and its validity, the proposed method was implemented using a real hydraulic excavator, and experiments using it were conducted.

3.2.1. Experimental setting

In this experiment, an operator operated the hydraulic excavator by remote control and excavated the ground. Obtained data from the fish-eye cameras was processed in real-time and an arbitrary viewpoint image was then generated. The experimental equipment and environment are shown in **Fig. 7**. Four fish-eye cameras were placed on the excavator body in different directions. The fish-eye cameras were Grasshopper3 GS3-U3-41C6C-C made by Point Grey Research. Their resolutions are 2048 × 2048 pixels, and their frame rate was set to 15 fps. Note that the operator watched the excavator directly and did not watch the arbitrary viewpoint image that was generated during the experiments.

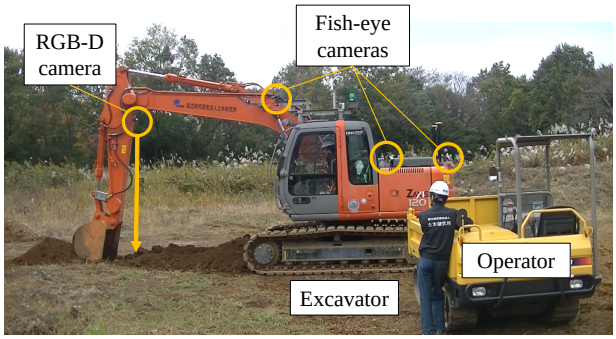

Fig. 7. : Experimental setting.

3.2.2. Results

Experimental results are shown in **Fig. 8**. **Fig. 8 (a)** is an arbitrary viewpoint image generated by the conventional method, and **Fig. 8 (b)** is an arbitrary viewpoint image generated by the proposed method. The arm of the excavator is in a wide area of the image in the conventional method, whereas it is removed and the hidden image behind it can be confirmed using the proposed method. An operator can omnidirectionally confirm the situation around an excavator assess any risks (such as collisions with obstacles), because the field of view is expanded using the proposed method.

3.3. Evaluation

The arbitrary viewpoint image generated by the proposed system is evaluated from two perspectives.

3.3.1. Real-time property of the image

As the objective of this study is to develop an image generation system that supports remote operation, the information that is obtained must be processed in real-time and presented to an operator. Therefore, the frame rate of arbitrary viewpoint images generated by the proposed system was measured. In this evaluation, the image drawing process was conducted 1000 times, and the average drawing time per one time was calculated from the required time for the whole process. The results show that the average drawing time for each process is 72.8 ms. This corresponds to 13.7 fps, and this system can provide images with a frame rate that is sufficient for remote operation [9]. Although the proposed system incorporates the process to determine the reflection of the arm and the visualization process of the ground shape, these do not prevent smooth remote operation.

(a) Conventional method.

(b) Proposed method.

Fig. 8. : Experimental result of image generation.

3.3.2. Expansion of view field by removal of occluded area

The expansion ratio of the view field of an arbitrary viewpoint image is validated by removing the reflection of the arm. In this evaluation, the ratio of the area on the dome-shaped model of the region in which the view field has been expanded by the proposed system to the region in which the arm had been reflected in the conventional system was calculated. In **Fig. 9**, the region in which the arm had been reflected in the conventional system is the region surrounded by the red frame, and the region in which the view field has been expanded by the proposed system is the region surrounded by the yellow frame. Furthermore, the region painted in black is the region in which images could not be obtained from any camera although the system determined that the arm is reflected. The ratio was calculated from the image from the fish-eye cameras and the joint angle data of the arm for approximately 30 s during one excavating motion. As a result, it has been confirmed that the view field expands in approximately 69.4% region within the region in which the arm had been reflected. This means that the proposed system can reduce approximately 69.4% of the occluded area that is generated by the conventional system and contribute to safer remote control of a hydraulic excavator. However, this

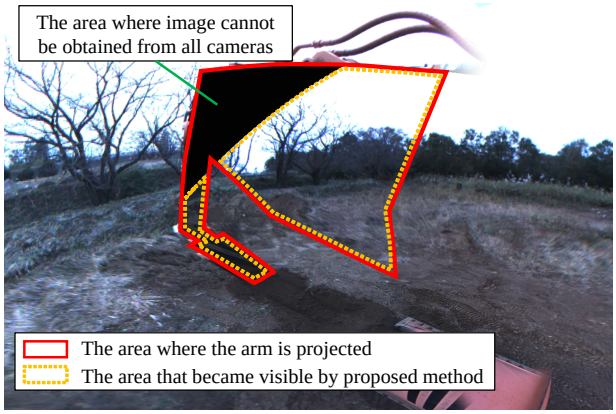

Fig. 9. : Expanded region of field of view.

method cannot provide images in the black region where images cannot be obtained from any camera, which is a limitation of this study. Although a simple solution is to install additional cameras to take images of the region, there are some cases where they cannot be installed owing to factors such as size, position, and processing capacity. One of the solutions to this problem is to use a complementary method using previous images. A method providing images passed through the arm by complementing from previous images according to the motion of the arm has already been proposed [10]. However, this method involves some degree of safety risk because it only uses previous images, and does not reflect the current situation. Therefore, it is necessary to use it in combination with another method, such as the detection and tracking of moving obstacles.

4. Presentation of 3D Information of Excavation Point

To solve the second issue, 3D information of the ground is integrated into the arbitrary viewpoint image generated by the method in chapter 3, and the ground shape is provided visually. In this chapter, the integration method, experiments using a real hydraulic excavator in which the proposed system is implemented, and an evaluation of the provided 3D information are described.

4.1. Method

For the proposed method, RGB-D sensor calibration and integration with arbitrary viewpoint images are required.

4.1.1. RGB-D sensor calibration

As is the case with the fish-eye cameras, because the RGB-D sensor that is used to measure the 3D shape is retrofitted cameras, their position and attitude with respect to the excavator are not known, and calibration is required. The calibration for the RGB-D sensor involves

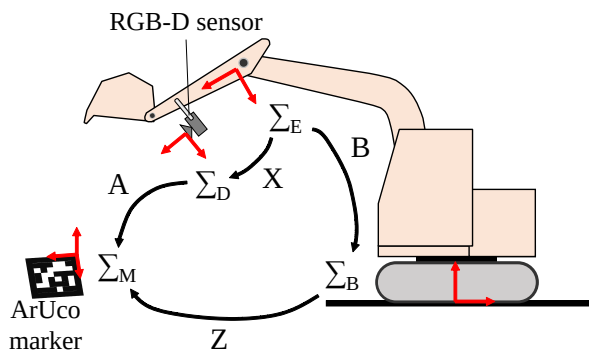

Fig. 10. : Coordinate systems for RGB-D sensor calibration.

deriving the transformation matrixes from the base coordinate system Σ_B of an excavator to the RGB-D sensor coordinate system Σ_D.

In this study, the RGB-D sensor is calibrated using an ArUco marker [11, 12] as a calibration pattern. Each coordinate system that is used here and the relationship between them are shown in **Fig. 10**. The transformation matrix **A** from the RGB-D sensor coordinate system Σ_D to the marker coordinate system Σ_M can be derived by detecting the ArUco marker that is fixed in the environment from an image of the RGB-D sensor by threshold processing and contour detection. Here, a finger coordinate system Σ_E is set on the arm on which the RGB-D sensor is placed. When the transformation matrix from the finger coordinate system Σ_E to the base coordinate system Σ_B of the excavator is defined as **B**, it can be derived by solving the forward kinematics from a 3D shape model of the arm and each joint angle data. Furthermore, the transformation matrix **X** from the finger coordinate system Σ_E to the RGB-D sensor coordinate system Σ_D, and the transformation matrix **Z** from the base coordinate system Σ_B to the marker coordinate system Σ_M are set as unknown parameters. Then, the following equation (11) is established.

$$\mathbf{AX} = \mathbf{ZB} \quad \ldots\ldots\ldots\ldots\ldots (11)$$

By moving only the arm to various positions and attitudes while the body of the excavator and the marker is fixed in the environment, and obtaining the numerous known combinations of the transformation matrixes $\mathbf{A}, \mathbf{B}$, these unknown transformation matrixes $\mathbf{X}, \mathbf{Z}$ can be derived analytically. In this study, the optimized calculation in which an analytic solution is set as an initial value is conducted by employing the method proposed by Dornaika et al. [13], and the transformation matrixes $\mathbf{X}, \mathbf{Z}$ are estimated simultaneously. From the above, the transformation matrix from Σ_B to Σ_D that is finally required can be derived.

4.1.2. Integration of 3D data

First, to integrate 3D information from the RGB-D sensor that is mounted on the arm to the arbitrary viewpoint image, a 3D point cloud is generated from depth information of each pixel of an RGB-D sensor image. The correspondence between a point ${}^{\mathrm{D}}\boldsymbol{p} = [x_{\mathrm{D}}, y_{\mathrm{D}}, z_{\mathrm{D}}]^\top$ on the RGB-D sensor coordinate system Σ_D and a point ${}^{\mathrm{d}}\boldsymbol{m} = [u_{\mathrm{d}}, v_{\mathrm{d}}]^\top$ on the RGB-D sensor image coordinate system Σ_d is represented as the following equation (12).

$$\begin{bmatrix} u_{\mathrm{d}} \\ v_{\mathrm{d}} \\ 1 \end{bmatrix} \simeq \begin{bmatrix} f_{\mathrm{x}} & 0 & c_{\mathrm{x}} & 0 \\ 0 & f_{\mathrm{y}} & c_{\mathrm{y}} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_{\mathrm{D}} \\ y_{\mathrm{D}} \\ z_{\mathrm{D}} \\ 1 \end{bmatrix} \quad \ldots (12)$$

f_{x}, f_{y}, c_{x}, and c_{y} in the equation are internal parameters of the sensor. In the case of the RGB-D sensor, depth information z_{D} along the z-axis can be obtained at a point ${}^{\mathrm{d}}\boldsymbol{m} = [u_{\mathrm{d}}, v_{\mathrm{d}}]^\top$ on the image. Therefore, a 3D point cloud can be generated by applying the equation (13) to all points on the image.

$$x_{\mathrm{D}} = \frac{u_{\mathrm{d}} - c_{\mathrm{x}}}{f/s_{\mathrm{x}}} z_{\mathrm{D}}, \quad y_{\mathrm{D}} = \frac{v_{\mathrm{d}} - c_{\mathrm{y}}}{f/s_{\mathrm{y}}} z_{\mathrm{D}} \quad \ldots\ldots (13)$$

The generated 3D point cloud having color information is transformed to the base coordinate Σ_B by using the results of the calibration, and it is integrated to the arbitrary viewpoint image.

Then, it is drawn after the region that is within the angle of view of the RGB-D image is removed from the arbitrary viewpoint image, meaning the information from the RGB-D image is represented preferentially in the region where the original arbitrary viewpoint image and the obtained point cloud overlaps.

However, it is difficult to understand the shape of the excavating point by only the 3D point cloud having the actual color. Therefore, to understand the shape of the excavating point more intuitively, this study proposes a visualization of the depth information of the ground using a mesh. Here, the depth information of the ground refers to the height or depth along the vertical direction based on the ground surface. The mesh is formed by connecting adjacent points of the sampled point cloud, and is drawn as a wireframe. It is rendered using a color that corresponds to the depth information. Using this mesh, the point cloud having original color information and the depth information can be recognized visually and simultaneously.

4.2. Experiments

To verify whether the proposed system can provide 3D information of the ground in real-time and its validity, the proposed method was applied to a real hydraulic excavator and experiments were conducted.

4.2.1. Experimental setting

In this experiment, an operator operated the hydraulic excavator by remote control and excavated the ground. Data obtained from the fish-eye cameras and the RGB-D sensor were processed in real-time, and an arbitrary viewpoint image that integrated 3D information of the ground was then generated. The experimental equipment and environment are shown in **Fig. 7**. In addition to four fish-eye

Fig. 11. : Experimental results of 3D data visualization: example of image.

cameras, an RGB-D sensor was equipped at the arm of the excavator to measure an excavation point. The RGB-D sensor that was used was RealSense D435, which was made by Intel, and the frame rate was set to 15 fps. Note that the operator watches the excavator directly, and did not watch the generated arbitrary viewpoint image during the experiments.

4.3. Results

Experimental results are shown in **Fig. 11** and **Fig. 12**. **Fig. 11** is an image when a viewpoint is set to watch an excavation point from the above. As shown in the figure, by combining a point cloud having an actual color and mesh visualizing depth information, the 3D shape of the excavating point can be understood intuitively. Moreover, **Fig. 12 (a)** is a side-view image that is generated by the conventional method, and **Fig. 12 (b)** is a side-view image that is generated by the proposed method. In the conventional method, the ground is represented as the flat plane even though it was already excavated. Conversely, in the proposed method, the ground shape at the excavating point can be understood. Because it is important for the excavation work performed by a hydraulic excavator to confirm whether the ground shape of the excavating point becomes close to the target shape, the proposed method is more practical than the conventional method. Furthermore, in **Fig. 11**, it can be seen that the shape of the excavated soil that is inside the bucket can also be measured. Although it is currently qualitative information, it is possible to know the amount of the excavated soil by comparing the shape of the empty bucket and the bucket containing the soil.

4.4. Evaluation

To increase the efficiency of excavation work by a hydraulic excavator, this system integrates 3D information from the RGB-D sensor equipped at the arm to the arbitrary viewpoint image. The position of the 3D data and the image may shift owing to errors such as the position and attitude of the RGB-D sensor estimated by calibration

(a) Conventional method.

(b) Proposed method.

Fig. 12. : Experimental results of 3D data visualization: comparison between conventional and proposed method.

and the joint angle of the arm. Therefore, this positional shift was evaluated quantitatively. In this evaluation, a chess pattern (the size of which is known) was placed on the ground, and the positional shift of the vertex of this pattern on a generated image was measured. The size of each side of the square of the used chess pattern is 0.13 m. As the proposed system approximates the ground using a flat plane, the pattern on the arbitrary viewpoint image and the image from the RGB-D sensor should be in agreement if the RGB-D sensor is ideally calibrated. As a result, the positional shift was 0.046 m when the 3D data of the RGB-D sensor were taken from a position 1.40 m away from the chess pattern. Therefore, the proposed system has a positional shift of approximately 3.3% to the dis-

tance from the RGB-D sensor to the target. It can be said that this positional shift is sufficiently small, and the required precision for excavation is satisfied.

5. Conclusion

In this study, we constructed an extended arbitrary viewpoint visualization system for the teleoperation of excavation work using a hydraulic excavator. We solved the two problems of occlusion by the arm and lack of 3D information about the excavating point. In the proposed method, it is determined whether the excavator's arm is included in the image, and the arbitrary viewpoint image is generated only from fish-eye camera images that do not capture the arm. As a result, the field of view of the generated arbitrary viewpoint image was expanded by 69.4% on average, and a visualization that contributes to safety in the teleoperation of excavation work was realized. Furthermore, the 3D information measured by an RGB-D sensor attached to the excavator's arm was synthesized using an arbitrary viewpoint image. The proposed method enables teleoperators to understand the 3D condition of the excavating point, which is necessary to determine the position of the bucket and was not possible with previous methods. Moreover, experiments and evaluations were conducted for each of the two solutions, and it was confirmed that an image with sufficient accuracy could be presented in real-time. Although this study focuses on hydraulic excavators, the proposed system can be applied to not only hydraulic excavators, but also whole machines with arms, and can provide arbitrary viewpoint images with a wide field of view to teleoperators in real-time in the same way.

Although the suitability of the proposed method was presented by the evaluation, further verification through actual field applications will be conducted. We plan to verify the improvement of the operability more quantitatively by using the proposed system with actual teleoperation experiments. In addition, as described in 4.4, it is important to reduce further occlusion in order to realize more safe and efficient teleoperation, and further consideration will be required to compensate for missing images that could not be solved by our proposed method. Furthermore, a manual or automatic control system to switch the viewpoint of the image according to the contents of work can be considered in the development of the proposed method. The realization of these advanced methods will lead to the development of a more practical visualization system, which will enable safer and more efficient teleoperation of construction machines and mobile robots.

Acknowledgements

This study was conducted with financial support from the Ministry of Land, Infrastructure, Transport, and Tourism of Japan. We would like to thank Editage (www.editage.com) for English language editing.

References:

[1] F. Matsuno and S. Tadokoro, "Rescue Robots and Systems in Japan," Proceedings of the 2004 IEEE International Conference on Robotics and Biomimetics, pp. 12–20, 2004.

[2] S. Kawatsuma, M. Fukushima, and T. Okada, "Emergency Response by Robotsto Fukushima-Daiichi Accident: Summary and Lessons Learned," Industrial Robot: An International Journal, vol. 39, no. 5, pp. 428–435, 2012.

[3] M. Moteki, K. Fujino, T. Ohtsuki, and T. Hashimoto, "Research on Visual Point of Operator in Remote Control of Construction Machinery," Proceedings of the 28th International Symposium on Automation and Robotics in Construction, pp. 532–537, 2010.

[4] S. Iwataki, H. Fujii, A. Moro, A. Yamashita, H. Asama, and H. Yoshinada, "Visualization of the Surrounding Environment and Operational Part in a 3DCG Model for the Teleoperation of Construction Machines," 2015 IEEE/SICE International Symposium on System Integration, pp. 81-87, 2015.

[5] W. Sun, S. Iwataki, R. Komatsu, H. Fujii, A. Yamashita, and H. Asama, "Simultaneous Tele-visualization of Construction Machine and Environment Using Body Mounted Cameras," Proceedings of the 2016 IEEE International Conference on Robotics and Biomimetics, pp. 382-387, 2016.

[6] M. Fuchida, S. Chikushi, A. Moro, A. Yamashita, and H. Asama: "Arbitrary Viewpoint Visualization for Teleoperation of Disaster Response Robots," Journal of Advanced Simulation in Science and Engineering, vol. 6, no. 1, pp. 249-259, 2019.

[7] D. Scaramuzza, A. Martinelli, and R. Siegwart, "A Flexible Technique for Accurate Omnidirectional Camera Calibration and Structure from Motion," Proceedings of the 2006 IEEE International Conference of Computer Vision Systems, p. 45, 2006.

[8] D. Scaramuzza, A. Martinelli, and R. Siegwart, "A Toolbox for Easily Calibrating Omnidirectional Cameras," Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 5695-5701, 2006.

[9] T. Sato, H. Fujii, A. Moro, K. Sugimoto, A. Nozue, Y. Mimura, K. Onata, A. Yamashita, and H. Asama, "Development of Bird's-Eye View System in Unmanned Construction," Transactions of the JSME (in Japanese), Vol. 81, Issue 823, pp. 14-31, 2015.

[10] T. Sato, A. Moro, A. Sugahara, T. Tasaki, A. Yamashita, and H. Asama, "Spatio-Temporal Bird's-Eye View Images Using Multiple Fish-eye Cameras," Proceedings of the 2013 IEEE/SICE International Symposium on System Integration, pp. 753-758, 2013.

[11] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic Generation and Detection of Highly Reliable Fiducial Markers Under Occlusion," Pattern Recognition, vol. 47, no. 6, pp. 2280-2292, 2014.

[12] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and R. Medina-Carnice, "Generation of Fiducial Marker Dictionaries Using Mixed Integer Linear Programming," Pattern Recognition, vol. 51, pp. 481-491, 2016.

[13] F. Dornaika and R. Horaud, "Simultaneous Robot-world and Hand-Eye Calibration," IEEE Transactions on Robotics and Automation, vol. 14, no. 4, pp. 617-622, 1998.

[14] R. Komatsu, H. Fujii, Y. Tamura, A. Yamashita, and H. Asama, "Free Viewpoint Image Generation System Using Fisheye Cameras and a Laser Rangefinder for Indoor Robot Teleoperation," ROBOMECH Journal, vol. 7, no. 1, pp. 1-10, 2020.

[15] F. Cosco, C. Garre, F. Bruno, M. Muzzupappa, and M.A. Otaduy, "Augmented Touch Without Visual Obtrusion," Proceedings of the IEEE International Symposium on Mixed and Augmented Reality 2009, pp. 99-102, 2009.