Simultaneous Tele-visualization of Construction Machine and Environment Using Body Mounted Cameras

Wei Sun*, Soichiro Iwataki*, Ren Komatsu*, Hiromitsu Fujii*, Atsushi Yamashita*, and Hajime Asama*

Abstract— **During the teleoperation of construction machines, it is highly needed to visualize the machine body itself. This is because the operator needs to perform a visual confirmation of the construction machine. In addition, information of crawler parts' shoe slip is indispensable. Therefore, not only the surrounding environment but also the vehicle body are needed to be visualized. This paper proposes a tele-visualization system for both the construction machine and the surrounding environment from an arbitrary viewpoint by using mounted cameras. In the proposed method, both the construction machine and the surrounding environment are expressed as three-dimensional mesh models. In order to capture the actual scenes of the construction machine and the surrounding environment, the positions and orientations of the mounted cameras are adjusted. Simultaneous tele-visualization is achieved by mapping the image data captured from the scenes to the correct positions on the mesh models, after estimated the correspondence between the camera images and the mesh models.**

I. INTRODUCTION

Construction machines are widely used in construction and disaster sites. However, manipulating a construction machine in construction or disaster sites, such as unmanned mines and landslide sites is usually dangerous for an operator [1]. In order to keep the operators' safety, teleoperation of construction machines is a feasible method [2], [3]. Since construction work sites are typically unique and full of uncertain situations, tele-visualization of the surrounding environment is indispensable. In addition, for efficient teleoperation of construction machines, the operator needs to grasp the condition of the vehicle body. In particular, visual confirmation of the construction machine itself is necessary for the manipulation. Furthermore, for example in the operation of a bulldozer, in order to adjust the load of the blade part to the most efficient load, the operator needs to grasp the shoe slip information of the crawler parts

Several visualization methods have been proposed in previous studies [4]–[7]. In [4], by mounting a panoramic camera onto the vehicle, visualization of the environment of the construction site was achieved. However, the operator has limited visibility while manipulating a construction machine from the cockpit [8]. In [5], visualization of the vehicle body using field cameras was proposed. However, unmanned mining construction work sites are usually spacious and dangerous. Hence, field cameras are not suitable. In [6], using four mounted out-ward facing cameras, visualization of the surrounding environment was well presented in a bird's-eye view. However, the actual scene of the vehicle body itself was not shown to the operator because the arrangement of the cameras was designed to capture the only surrounding environment. In [7], both the surrounding environment and the blade part of the construction machine were made possible to be visualized. Visualization method of the surrounding environment was the same as [6]. An in-ward facing camera was used to visualize the blade part of the construction machine. However, in order to visualize the whole vehicle body by the method mentioned in [7], a large number of in-ward facing cameras will be needed, because every operation part such as the crawlers will need an in-ward facing camera to be visualized.

Thus, the purpose of this study is to develop a simultaneous tele-visualization system for both the body of the construction machine and the surrounding environment. The main approach of this study is to project the actual scene image captured from the mounted cameras onto the screens. The screens of the construction machine and the surrounding environment are prepared, respectively, as three-dimensional mesh models. By mapping the textures obtained from the cameras onto the three-dimensional mesh models, the simultaneous visualization of both the scenes can be achieved.

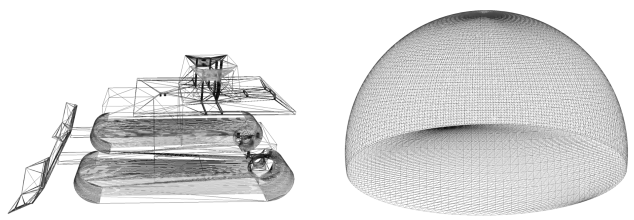
II. PROPOSED METHOD

In order to simultaneously visualize both the construction machine and the surrounding environment, a method that consists of four steps is proposed. Firstly, three-dimensional mesh models of the construction machine and the surrounding environment are built. Secondly, in order to obtain images of the actual scenes of the construction machines and the surrounding environment simultaneously, positions and orientations of the cameras are adjusted. Thirdly, the correspondence between the camera images and the three-dimensional mesh models is estimated. Finally, images of the actual scenes are mapped to both three-dimensional models. In this manner, simultaneously tele-visualization system is proposed.

A. Three-dimensional mesh models of the construction machine and the surrounding environment

As the screens for projecting actual scene images, three-dimensional mesh models of the construction machine and the surrounding environment are needed. In this study, the construction machine mesh model and the surrounding environment mesh model are built in different ways. For the construction machine, because the CAD data is available, it is possible to build an accurate mesh model in proper precision

*W. Sun, S. Iwataki, R. Komatsu, H. Fujii, A. Yamashita, and H. Asama are with the Department of Precision Engineering, Faculty of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. {sun, iwataki, komatsu, fujii, yamashita, asama}@robot.t.u-tokyo.ac.jp



(a) (b)

Fig. 1: (a) The mesh model of the construction machine is built by CAD data. Thus the mesh model of the construction machine can be built in proper precision. (b) The surrounding environment is approximated as a semi-sphere mesh model, because the surrounding environment of construction work sites is usually spacious and unable to be measured.

(Fig. 1(a)). The surrounding environment of construction work sites is usually spacious and difficult to measure. Therefore, in this study, the surrounding environment is approximated as a semi-sphere mesh model (Fig. 1(b)). The radius of the semi-sphere can be adjusted according to the region that operators need to look over in the actual site. Thus, the mesh model of the construction machine and the surrounding environment are built.

In order to present the construction machine in the center of the surrounding environment. in this study, the origins of the construction machine mesh model coordinate system and the surrounding environment mesh model coordinate system were matched.

B. Positions and orientations of cameras

In order to look over the surround of the machine, the field of view of 360° is necessary. In this study, the wide field of view is obtained by four fish-eye cameras that are mounted on all sides of the construction machine. A fish-eye camera is a wide angle camera, which has an over 180° angle of view. Moreover, in order to obtain both actual scene images of the construction machine and the surrounding environment, the orientations of fish-eye cameras are adjusted. In this study, the fish-eye cameras are adjusted to downward directions as shown in Fig. 2. The visible region of left side mounted fish-eye camera is the non-shadow part in Fig. 2. Also, the orientation of the fish-eye camera is shown as the red arrow.

C. Correspondence between the camera images and the mesh models

In order to obtain the correspondence between the camera images and the mesh models, calibrations between cameras and mesh models are needed. In this study, the calibration has been divided into estimating intrinsic parameters of the fish-eye cameras and estimating extrinsic parameters of the fish-eye cameras.

1) Intrinsic parameters: In this study, the intrinsic parameters of fish-eye cameras are estimated by the method in [9]. A given three-dimensional point (a,b,c) in the world coordinate system is projected as a direction vector $\boldsymbol{p}$ in the

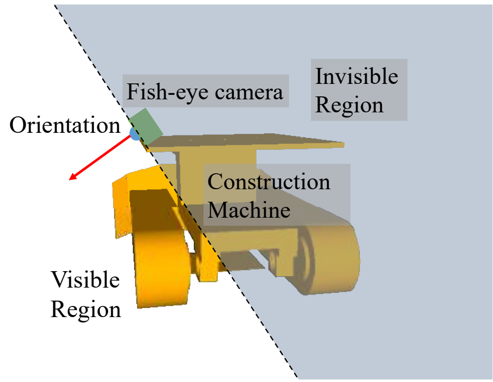


Fig. 2: In order to obtain the actual scene images of both the construction machine and the surrounding environment, fish-eye cameras are adjusted to downward directions. The visible region of left side mounted fish-eye camera is the non-shadow part after the installation angle of fish-eye camera is adjusted to a lower diagonal direction. The red arrow shows the the orientation of the fish-eye camera.

fish-eye camera coordinate system on the spherical camera model (Fig. 3). This is further projected to coordinates (u,v) in the fish-eye image as shown in Eq. (1).

$$\boldsymbol{p} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} u \\ v \\ f(\rho) \end{bmatrix}. \quad (1)$$

The relation between the spherical coordinate (x,y,z) and (u,v) is obtained as a 4^{th} order polynomial (Eq. (2)).

$$f(\rho) = a_0 + a_1\rho + a_2\rho^2 + a_3\rho^3 + a_4\rho^4, \quad (2)$$

$$\rho = \sqrt{u^2+v^2}. \quad (3)$$

The coefficients a_i $(i = 0,\ 1,\ 2,\ 3,\ 4)$ of the polynomial are intrinsic parameters.

2) Extrinsic parameters: Extrinsic parameters can be estimated after the intrinsic parameters are known. In order to estimate extrinsic parameters, the world coordinate system is

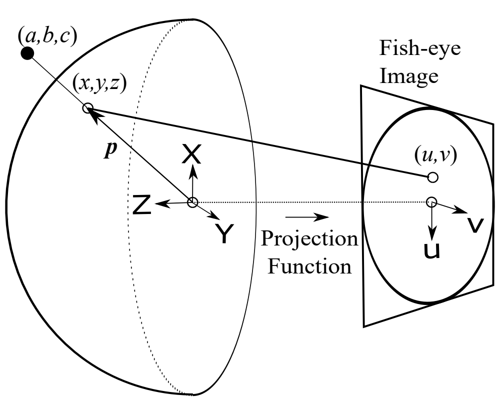


Fig. 3: A three-dimensional point (a,b,c) in the world coordinate system is projected as a direction vector $\boldsymbol{p}$ in local fish-eye coordinate system on the spherical camera model. Also, it is projected to coordinate (u,v) in the fish-eye image.

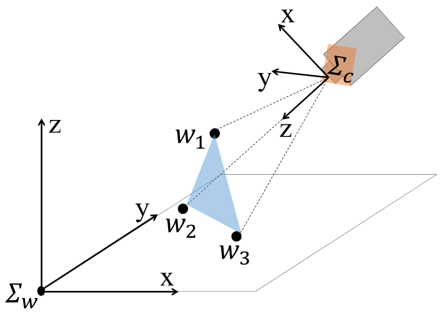


Fig. 4: The world coordinate system is denoted as Σ_w. The local fish-eye camera coordinate system is denoted as Σ_c. The extrinsic parameters between the local fish-eye camera coordinate system and the world coordinate system can be estimated by using three known points w_1, w_2, w_3.

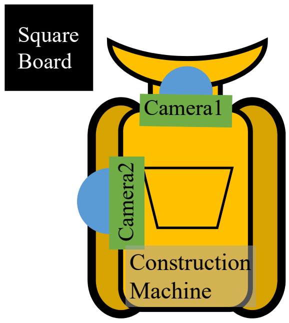


Fig. 5: Square boards are used for estimating the extrinsic parameters between the local fish-eye camera coordinate system and the surrounding environment mesh model coordinate system. In order to estimate the extrinsic parameters between the local fish-eye camera coordinate systems of camera 1 and camera 2, a square board is placed in the common visible region of camera 1 and camera 2.

denoted as Σ_w. The local fish-eye camera coordinate system is denoted as Σ_c. As mentioned in the previous study [10], it is possible to estimate the extrinsic parameters by three known points w_1, w_2, w_3 in the world coordinate system (Fig. 4). In this study, the extrinsic parameters between the local fish-eye camera coordinate system and the construction machine mesh model coordinate system, and the extrinsic parameters between the local fish-eye camera coordinate system and the surrounding environment mesh model coordinate system are estimated.

Because the shape of the construction machine is known as the CAD model, the correspond points of specified image points on the construction machine mesh model are possible to be obtained. Therefore, the extrinsic parameters between the construction machine mesh model coordinate system and the local fish-eye camera coordinate system can be estimated by the method mentioned in [10].

In order to estimate the extrinsic parameters between the local fish-eye camera coordinate system and the surrounding environment mesh model coordinate system, square boards are used. Estimation of the extrinsic parameters between each local fish-eye camera coordinate system are needed in this study. As already mentioned, the reference points need to be captured by at least two cameras. Therefore, in this study, the square board whose four vertexes of the corners are referential points is placed in the visible region of each two fish-eye cameras. For example, Fig. 5 shows how the square board is placed in the common visible region from both front side and left side fish-eye cameras. By using the square board, both extrinsic parameters between the local fish-eye camera coordinate system and the surrounding environment mesh model coordinate system and the extrinsic parameters between each local fish-eye camera coordinate system are estimated.

Thus, the correspondence between the camera images and both mesh models can be estimated.

D. Texture mapping

In order to simultaneously visualize both the construction machine and the surrounding environment, the images of the actual scenes are mapped to the corresponding position in the mesh models. Moreover, in order to show an easy-to-understand visualization, the parts of the construction machine mesh model of which the cameras cannot obtain the actual scene images are colored in the original color of those parts in advance.

III. Simultaneous Tele-visualization Experiment Using Proposed Method

A. Experimental settings

Experiment was performed with a bulldozer model shown in Fig. 6. In order to verify that the parts of the machine

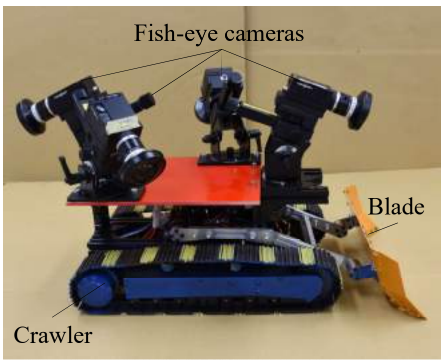


Fig. 6: Experiment was performed with this bulldozer model. The color of the bulldozer model was changed in order to be exhibited clearly. The upper part of the bulldozer was painted in red, the blade part was painted in orange, and the crawler parts were painted in blue and yellow. This bulldozer model is powered by AC adapters of 5 V and 12 V, and controlled by a USB gamepad.

were shown on the appropriate parts on the mesh model, the model was painted in multicolor as shown in Fig. 6. The upper part of the bulldozer model was painted in red, the blade part was painted in orange, and the crawler parts were painted in blue and yellow. This bulldozer model is powered by AC adapters of 5 V and 12 V, and controlled by a USB gamepad.

The surrounding environment of experiment was approximated as a semi-sphere mesh model with 1 m radius. The mesh model of the bulldozer model was built by CAD software from previously known CAD data. In order to integrate the coordinate systems of the mesh models in a common coordinate system, the origins of the bulldozer model mesh model coordinate system and the surrounding environment mesh model coordinate system were matched. In order to observe the 360° images of the actual scenes, four fish-eye cameras are mounted onto the bulldozer model as shown in Fig. 6. The fish-eye cameras were placed in the front, upper left, back, and lower right side of the bulldozer model as shown in Fig. 7. In addition, the fish-eye cameras are numbered as C_i $(i = 1,2,3,4)$ in a counterclockwise order. The colored shadows: blue, orange, green, and gray show the visible region of each fish-eye camera. The common visible regions of every pair of fish-eye cameras are labeled as C_i,C_j $(i, j = 1,2,3,4)$. The orientations of the fish-eye cameras are shown as red arrows. In order to obtain real time images of both the self body of the construction machine and the surrounding environment by fish-eye cameras, the left side and right side installation angles of the fish-eye cameras were adjusted to 40° below the horizon. In addition, the front side and back side installation angles of the fish-eye cameras were adjusted to 20° below the horizon. Thus, both the bulldozer model and the surrounding environment were able to be visualized simultaneously. The positions and orientations of the mounted fish-eye cameras are shown in Fig. 8. In addition, in order to estimate the extrinsic parameters, four square boards were placed in the common visible regions. The four square boards are numbered as B_i $(i = 1,2,3,4)$ in a counter-clockwise order

In order to visualize the bulldozer model and the surrounding environment simultaneously, calibration was preformed by the method mentioned in Section II-C. The upper part of the bulldozer model of which cameras could not obtain the actual scene images, was colored in red in advance.

B. Simultaneous tele-visualization experiment

The simultaneous tele-visualization experiment was performed with the experiment set up mentioned above. The bulldozer model was operated by a gamepad during the experiment. Fish-eye cameras were set to capture the images of the actual scenes at 24 fps. Meanwhile, the images were mapped to the mesh models. The result of the texture mapping was presented on a monitor of a computer at 24 fps in the same time. Thus, operation of the bulldozer model by observing the monitor of a computer was possible.

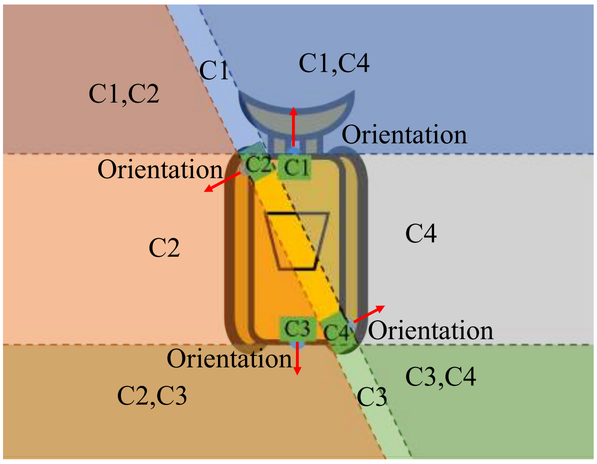


Fig. 7: Positions and orientations of C_i $(i = 1,2,3,4)$. Four mounted fish-eye cameras numbered in the order of counterclockwise. The color shadows: blue, orange, green, and gray show the visible region of each fish-eye camera. The common visible region of every two fish-eye cameras labeled in C_i,C_j $(i, j = 1,2,3,4)$. The orientations of the fish-eye cameras are shown as the red arrows.

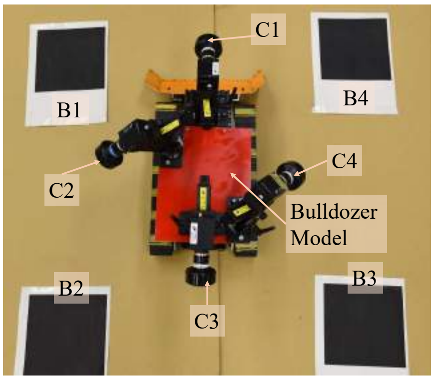
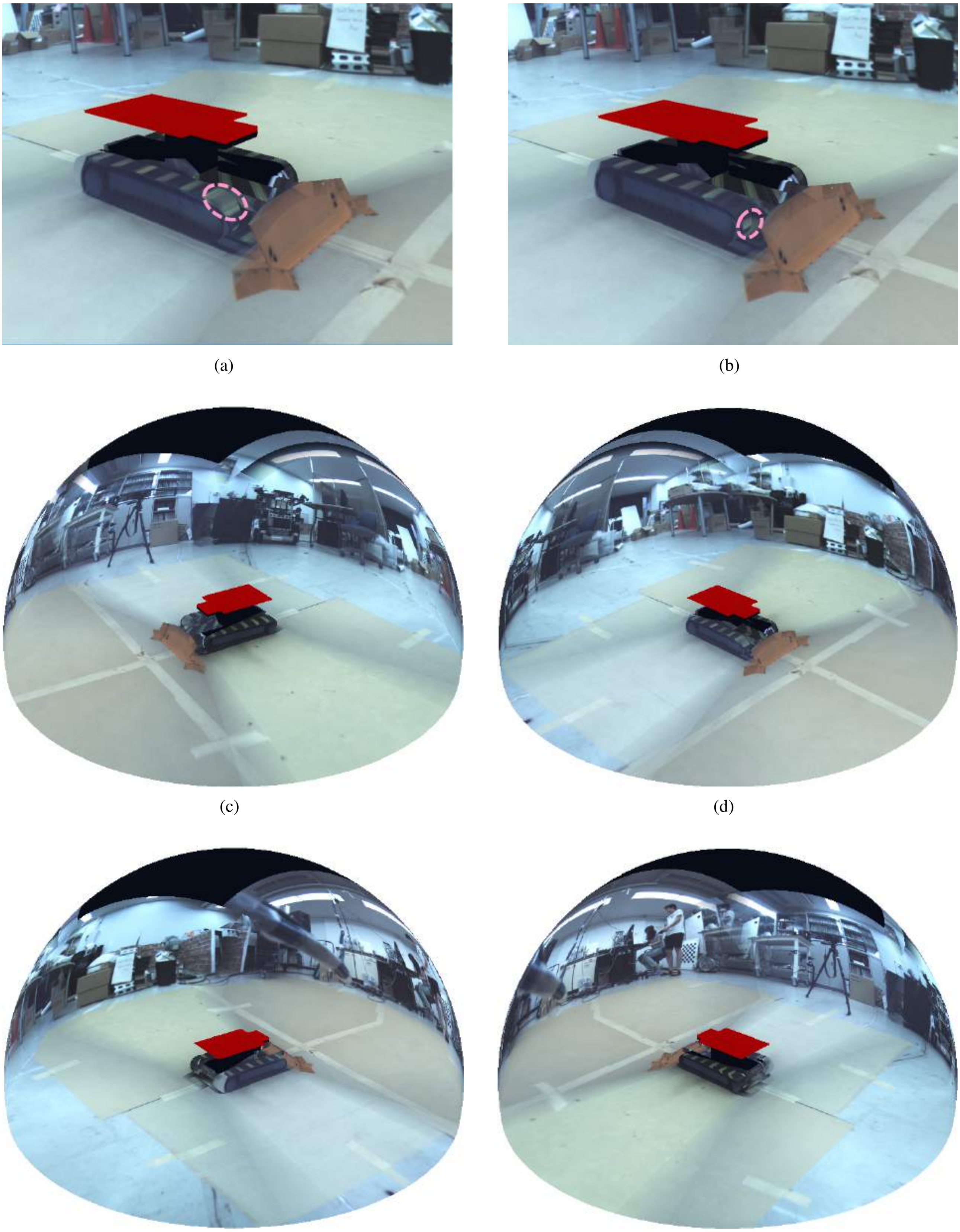


Fig. 8: Installation angle of fish-eye cameras. The left side and right side installation angle of fish-eye camera were adjusted to 45° lower than horizontal direction. The front side and back side installation angle of fish-eye camera adjusted to 10° lower than horizontal direction. Four square boards B_i $(i = 1,2,3,4)$ were placed in common visible region of every two fish-eye cameras in a counter-clockwise order.

C. Experimental result

Experiment result is shown in Fig. 9. The bulldozer model crawler part and four angular views of the tele-visualization are shown in Fig. 9. Since the experiment result is presented as an on-line video observed from the monitor, the rotation of the bulldozer model crawler part and the movement of the bulldozer model can be confirmed from the on-line video. The relation between the bulldozer model and the surrounding environment is possible to be confirmed as well. Figures 9(a)–(b) show two continuous frames of the on-line video. The striped pattern which was marked by the



(a) (b) (c) (d) (e) (f)

Fig. 9: The bulldozer model crawler parts and four angular views of the tele-visualization are shown. Figures (a) and (b) show 2 continuous frames of the on-line video. The striped pattern which was marked by the pink dashed line shows the rotation of the bulldozer model crawler part. Both the bulldozer model and the surrounding environment are visualized simultaneously as shown in figures (c)–(f).

pink dashed line is shown in different positions according to Figures 9(a)–(b), because the crawlers' rotation. According to Figures 9(c)–(f), it was possible to confirm that, the calibration worked properly and that the correspondence between the camera images and the mesh model was correct. The actual scene images were mapped to the mesh models successfully. Moreover, it was possible to visualize both the bulldozer model and the surrounding environment in an arbitrary view. By using the method proposed above, operators may grasp the information of the shoe slip of the crawler parts and may perform the visual confirmation by the tele-visualization.

IV. CONCLUSIONS

In this study, we developed a system for simultaneous tele-visualization of both the construction machine and the surrounding environment.

The adequacy of the system was confirmed by a simultaneous tele-visualization experiment. The result of the experiment showed the change of the surrounding environment during the operation of the bulldozer model clearly. In addition, the rotation of the bulldozer model crawler parts was also visualized during the experiment. In comparison with the previous study [7], instead of using five cameras, four mounted fish-eye cameras were used. By adjusting the position and orientations of the mounted cameras, simultaneous tele-visualization of both the construction machines and the surrounding environment was possible, instead of visualizing the environment and only the blade part of the bulldozer. The proposed method is possible to be used in every type of construction machine. Thus, the system for simultaneous tele-visualization was developed successfully.

For future work, optimal positions and orientations of the mounted fish-eye cameras will be considered. In addition, because of the movements of the operation parts during manipulation, it is necessary to change the mesh model accordingly.

REFERENCES

[1] M. McCann, "Heavy equipment and truck-related deaths on excavation work sites," *Journal of Safety Research*, Vol. 37, No. 5, pp. 511–517, 2006.

[2] K. Bayne, R. Parker, "The introduction of robotics for New Zealand forestry operations: Forest sector employee perceptions and implications," *Technology in Society*, Vol. 34, No. 2, pp. 138–148, 2012.

[3] T. Hirabayashi, J. Akizono, T. Yamamoto, H. Sakai, H. Yano, "Teleoperation of construction machines with haptic information for underwater applications," *Automation in Construction*, Vol. 15, No. 5, pp. 563–570, 2006.

[4] C. James, T. Bednarz, K. Haustein, L. Alem, C. Caris, A. Castleden, "Tele-operation of a mobile mining robot using a panoramic display: an exploration of operators sense of presence," *Proceeding of the 2011 IEEE International Conference on Automation Science Engineering*, pp. 279–284, 2011.

[5] F. Okura, Y. Ueda, T. Sato, N. Yokoya, "Teleoperation of mobile robots by generating augmented free-viewpoint images," *Proceedings of the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 665–671, 2013.

[6] T. Sato, A. Moro, A. Sugahara, T. Tasaki, A. Yamashita, 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.

[7] S. Iwataki, H. Fujii, A. Moro, A. Yamashita, H. Asama, H. Yoshinada, "Visualization of the surrounding environment and operational part in a 3DCG model for the teleoperation of construction machines," *Proceedings of the 2015 IEEE/SICE International Symposium on System Integration*, pp. 81–87, 2015.

[8] J. Teizer, Ben S. Allrea, U. Mantripragada, "Automating the blind spot measurement of construction equipment," *Automation in Construction*, Vol. 19, No. 4, pp. 491–501, 2010.

[9] D. Scaramuzza, A. Martinelli, 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.

[10] R. Haralick, C. Lee, K. Ottenberg, M. Nolle, "Review and analysis of solution of the three point perspective pose estimation problem," *International Journal of Computer Vision*, Vol. 13, No. 3, pp. 331–356, 1994.