

Visualization of the Surrounding Environment and Operational Part in a 3DCG Model for the Teleoperation of Construction Machines

Soichiro Iwataki¹, Hiromitsu Fujii¹, Alessandro Moro^{1,2},
Atsushi Yamashita¹, Hajime Asama¹, and Hiroshi Yoshinada³

Abstract—During the teleoperation of construction machines and robots, it is essential that the surrounding environment, operational part, and work object are visually presented to the operator. In this paper, we present a novel image representation method that combines these three aspects from an arbitrary point of view using multiple cameras. The image generated by the proposed method is compared to those produced by a conventional method.

I. INTRODUCTION

Recently, technology for the teleoperation of construction machines and robots has attracted considerable attention. For example, construction works where there is a risk of debris flow are often carried out by teleoperated construction machines that are controlled by a remote operator. Teleoperated robots are also used for remote investigation such as in the nuclear power plant which is destroyed by The Great East Japan Earthquake [1], [2].

However, it is known that teleoperation performance is affected by limited visibility such as a narrow field of view or multiple windows [3], [4]. Therefore, for safe and effective teleoperation, it is essential to present the surrounding environment and work objects in appropriate form.

To present the situation of the surrounding environment and work objects to the operator, various image representation methods have been proposed [5]–[12].

However, in construction machines such as bulldozers, whose operational parts are in front of them, it is impossible to observe both the operational parts and work objects. Figure 1 illustrates such a situation. The operational part of the bulldozer (blade) is shown by the blue balloon in Fig. 1. The red part represents the blind spot of an operator in the cabin. Therefore, image presentation methods [5] that present

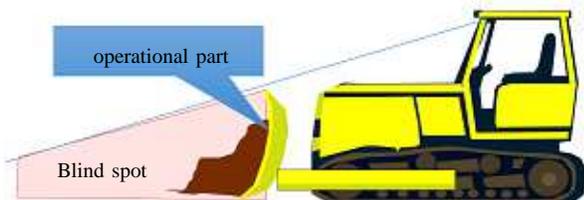


Fig. 1. Blind spot from cockpit

an image as if acquired from the cockpit are unsuitable for the teleoperation of bulldozers. Hashimoto et al. [8], Okura et al. [9], and Kelly et al. [13] have proposed methods that present images as if the operator is observing the machine directly. However, the image presentation method used in [8] requires field cameras. Therefore, it is not suitable for wide areas such as construction sites or disaster zones. Moreover, construction and debris removal tasks change the features of the terrain, unlike investigation tasks which basically do not modify the landscape. Therefore, methods [9], [13] that utilize and gather depth data into a 3D map is unfeasible. For same reason, SLAM-based method [6], [7] are also infeasible.

Other image presentation methods use images acquired from cameras [10], [11]. These methods do not rely on a global map, and are therefore feasible under conditions in which terrain features may be changed.

These approaches present images as if the operator is observing the machine and the surrounding environment from above. Sato [11] reported that operators can easily recognize the surrounding environment of a construction machine from a single image, and that the efficiency of driving tasks in the presence of obstacles is increased by these overhead images compared to images from the cockpit.

However, these presentation methods are mainly focused on the surrounding environment. Therefore, they cannot visualize the operational part and work objects, because only images from above the construction machine are presented.

Thus, teleoperate construction machines whose operational part and work objects cannot be observed from the cockpit, it is desirable to present images from arbitrary viewpoints without 3D mapping, such as by point cloud registration.

This research has developed an image presentation method that enable operator to easily and simultaneously recognize the surrounding environment, operational part, and work objects .

Our image presentation system uses simple geometric shapes and 3DCG model. The presented image is generated by mapping the images acquired by cameras mounted on the chassis of construction machines or robots to the 3DCG model.

II. PROPOSED PRESENTATION METHOD

The proposed method presents images of the surrounding environment which acquired by outward-facing cameras, mapping these images to hemisphere shape 3DCG model. The operational part and work objects which can not be

¹Graduate School of Engineering, The University of Tokyo, Tokyo, Japan

²Ritecs, Tokyo, Japan

³Komatsu Joint Research Laboratory, Graduate School of Engineering, Osaka University, Osaka, Japan

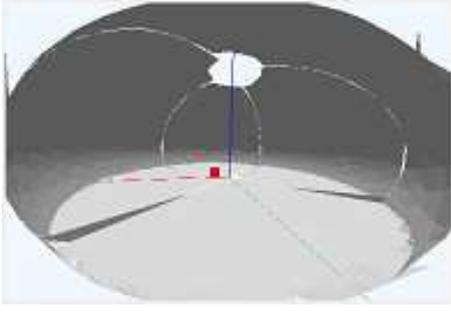


Fig. 2. Sphere mesh for presentation of surrounding environment

observed by these outward cameras are observed by an inward-facing camera and mapped to 3DCG model of the operational part. These 3DCG models consist of triangles. The combination of these images results enables the operator to recognize the surrounding environment, operational part, and work objects.

A. Presentation of surrounding environment

For the safe operation of teleoperated machinery, it is essential that the operator can see the surrounding environment. To visualize the surrounding area with as few cameras as possible, fisheye cameras are used.

Images of the surrounding environment are acquired by four fisheye cameras mounted on the machine. These images are integrated by the method based on [11], and are then projected onto a hemispheric triangle mesh that surrounds the 3D model of the machine.

Figure 2 shows the triangle mesh used for texture mapping. Figure 3 shows some example of images acquired by the fisheye cameras. The hemisphere shaped 3DCG model of surrounding environment is obtained in such a way that these images can be projected onto the mesh structure.

B. Presentation of operational part and work object

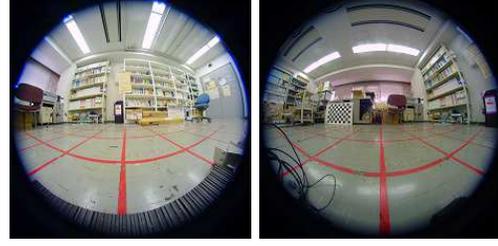
It is impossible to visualize the operational part and work objects using fisheye images. For example, Fig. 3(d) shows the image acquired by the front fisheye camera. It is impossible to show the operational part, work objects, and part of the surrounding environment using this image.

Therefore, the unobservable region of the surrounding environment, operational part, and work objects are visualized by projecting the image acquired by the camera above the operational part to 3DCG model of the operational part which consist of triangle meshes.

The cameras are positioned in such a way that four fisheye cameras can observe the surrounding environment and one camera can observe the operational part and the ground in front of the operational part.

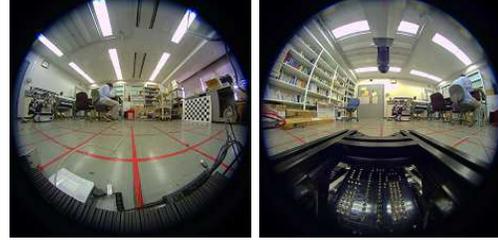
To present a fine image of the operational part and the work object, normal camera is used to observe operational part and work object, instead of fisheye camera.

Assume that a real camera and a virtual camera are represented by a perspective camera.



(a) Left fisheye camera image

(b) Back fisheye camera image



(c) Right fisheye camera image

(d) Front fisheye camera image

Fig. 3. Input images used to present surrounding environment



Fig. 4. Presented image of surrounding environment

In other words, when a 3D spatial position $\mathbf{x} = (x, y, z)^T$ is allocated at coordinate $\mathbf{m} = (u, v)^T$ in an image, the homogeneous coordinates of the points $\tilde{\mathbf{x}} = (x, y, z, 1)^T$, $\tilde{\mathbf{m}} = (u, v, 1)^T$, and the projection matrix \mathbf{P} are formulated by the following equation:

$$\tilde{\mathbf{m}} \sim \mathbf{P}\tilde{\mathbf{x}}. \quad (1)$$

In this way, images of the operational part acquired by the camera and the operational part in the 3DCG model are associated by a projective transformation.

Additionally, the projective transformation matrix \mathbf{P} can be obtained from six pairs of correlation points. Vector \mathbf{p} is defined as following equation using the elements of matrix $\mathbf{P} = (p_{ij})$.

$$\mathbf{p} = (p_{11}, p_{12}, p_{13}, p_{14}, p_{21}, p_{22}, p_{23}, p_{24}, p_{31}, p_{32}, p_{33}, p_{34})^T \quad (2)$$

In this case, the following two equations are obtained from the pair $(x_i, y_i, z_i)^T$, (u_i, v_i) :

$$\begin{aligned} x_i p_{11} + y_i p_{12} + z_i p_{13} + p_{14} \\ -x_i u_i p_{31} - y_i u_i p_{32} - z_i u_i p_{33} - p_{34} = 0 \end{aligned} \quad (3)$$

$$\begin{aligned} x_i p_{21} + y_i p_{22} + z_i p_{23} + p_{24} \\ -x_i v_i p_{31} - y_i v_i p_{32} - z_i v_i p_{33} - p_{34} = 0. \end{aligned} \quad (4)$$

Twelve equations are obtained from any given six pairs of points. These equations are expressed as follows equation using a 12×12 matrix \mathbf{B} and 12-dimensional vector \mathbf{p} .

$$\mathbf{B}\mathbf{p} = \mathbf{0}. \quad (5)$$

To determine \mathbf{p} satisfying Eq. 5, the following optimization process is carried out.

$$\text{minimize } \|\mathbf{B}\mathbf{p}\|^2 \quad (6)$$

$$\text{s.t. } \|\mathbf{p}\|^2 = 1. \quad (7)$$

This equation is solved using Lagrange multiplier method.

$$L = (\mathbf{B}\mathbf{p})^T(\mathbf{B}\mathbf{p}) - \lambda(\mathbf{p}^T\mathbf{p} - 1) \quad (8)$$

$$\frac{\partial L}{\partial \mathbf{p}} = 2(\mathbf{B}^T\mathbf{B}\mathbf{p} - \lambda\mathbf{p}) = 0. \quad (9)$$

The vector \mathbf{p} that satisfies this equation is the eigenvector corresponding to minimum eigenvalue of matrix $\mathbf{B}^T\mathbf{B}$.

In this way, \mathbf{P} can be found.

In the proposed method, camera mounting errors when the system starts up mean that \mathbf{P} , which represents the projective transformation to the operational part and the area in front of the machine, is unknown. Thus \mathbf{P} is calculated when the system is initialized using points obtained from an image acquired by camera above the operational part and pre-defined points in the 3DCG model of operational part.

The red points in Fig. 8 show these pre-defined calibration points.

Corner points in 3DCG model of operational part are selected as the point used to calculate \mathbf{P} . Thus, in the proposed method, the 3DCG model of the operational part must contain six different corner points. The Points $\mathbf{x}_i = (x_i, y_i, z_i)^T$ are given in this way.

In contrast, points $\mathbf{m}_i = (u_i, v_i)^T$ corresponding to $\mathbf{x}_i = (x_i, y_i, z_i)^T$ in input image are given manually. \mathbf{P} is calculated from these six point pair,

Using the \mathbf{P} obtained by initializing process, the other points $\mathbf{m}_i = (u_i, v_i)^T$ corresponding to $\mathbf{x}_i = (x_i, y_i, z_i)^T$ in the 3DCG operational part model are calculated.

In this way, points corresponding to the vertices of each triangle in the 3DCG model are calculated.

In a real system, the input image of the operational part is of finite size. The projected point of the 3DCG triangle mesh might be located outside the range of the input image. In this case, it is necessary to modify the triangle mesh so that the projected vertex points are within range of input image. This modification process is carried out as follows. The triangle which has vertices $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ is projected to points $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3$ in the input image, respectively, by the projection matrix \mathbf{P} . If all of the points $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3$ are



Fig. 5. Input image of the operational part and work object

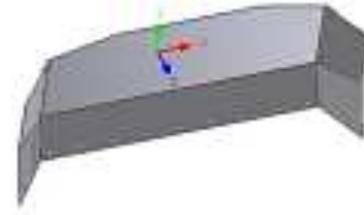


Fig. 6. The operational part

outside the image, or if at least one of the edges of triangle is shorter than some pre-configured threshold length, the triangle is removed from the mesh.

If at least one of the points $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3$ is within range, the triangle is broken down into four small triangles. We compute point \mathbf{v}_a as the midpoint of edge $\mathbf{v}_1 - \mathbf{v}_2$, point \mathbf{v}_b as the midpoint of edge $\mathbf{v}_2 - \mathbf{v}_3$, point \mathbf{v}_c as the midpoint of edge $\mathbf{v}_3 - \mathbf{v}_1$.

The four small triangles are then $(\mathbf{v}_1, \mathbf{v}_a, \mathbf{v}_c)$, $(\mathbf{v}_a, \mathbf{v}_2, \mathbf{v}_b)$, $(\mathbf{v}_b, \mathbf{v}_3, \mathbf{v}_1)$, $(\mathbf{v}_a, \mathbf{v}_b, \mathbf{v}_c)$. These four triangles are added to the mesh.

This modification process is carried out recursively until all projected points of triangle is in range of the input image.

Once the initialization process has been completed, the projective transformation is carried out using the invariant \mathbf{P} while the system is operating because the relative position of cameras respect to machine body is invariant. In the proposed method, no additional object, such as chess board pattern, is necessary to compute \mathbf{P} .

Using the matrix \mathbf{P} obtained by the initialization process, the points $\mathbf{m}_i = (u_i, v_i)^T$ corresponding to $\mathbf{x}_i = (x_i, y_i, z_i)^T$ in the 3DCG operational part model are calculated.

For example, Fig. 5 shows the operational part and the plane of the ground in front of the operational part. Figures 6 and 7 represent the 3DCG model of operational part and work object used in presentation process.

The operational part and the ground plane are presented in 3DCG model (see Fig. 9) via texture mapping to 3DCG model from Fig. 5.



Fig. 7. Work object

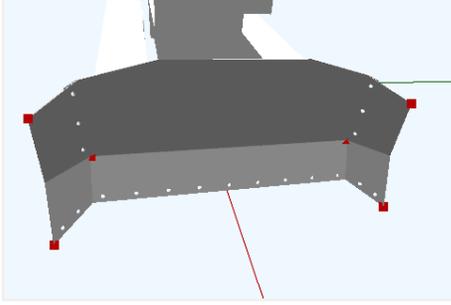


Fig. 8. Pre-defined calibration point of the operational part.

III. TELEOPERATION EXPERIMENT USING THE PROPOSED METHOD

To examine the effectiveness of the proposed method, teleoperation experiment was carried out.

The experiment was conducted using a small bulldozer equipped with a system of cameras.

A. Experiment system

Figure 10 shows the bulldozer used for experiment.

This model is externally powered by 5 V and 12 V AC adapters, and is controlled by a USB gamepad connected to a laptop PC.

In Fig. 10, the fisheye cameras are highlighted by yellow circles. These four fisheye cameras are mounted on the front, back, left, and right of the bulldozer, and acquire images of the surrounding environment.

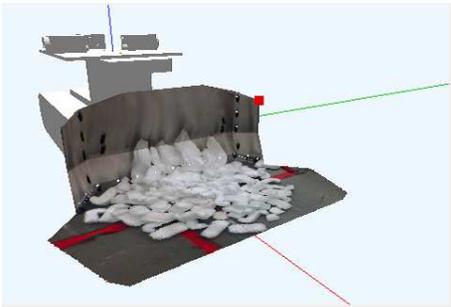


Fig. 9. The example of texture mapping to the operational part and work object.

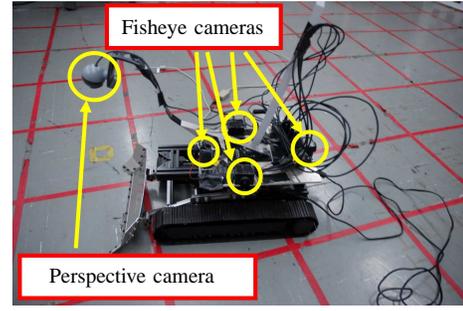


Fig. 10. Model bulldozer used in the experiment

In addition to these cameras, another normal camera is mounted in such a way as to acquire images of operational part and work objects. The cameras are connected to the laptop PC via a USB2.0 hub.

B. Comparison between the presented image and real image

To evaluate the geometrical appropriateness of the proposed method, the presented image was compared to a real image.

Under the assumption that the real world is observed by a single perspective camera, lines in the real world are projected to lines in the obtained image and the relation between these lines are preserved. It is desirable that the image produced by the proposed method satisfies these relationships.

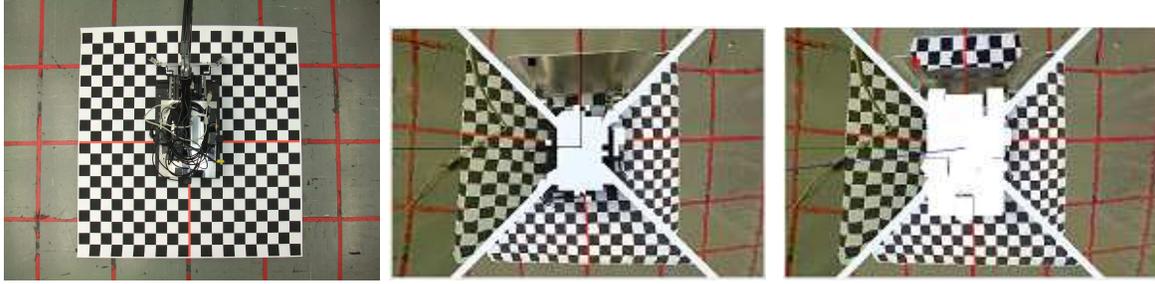
Figure 11 shows two different picture of the same scene. One is the image generated by the method of Sato et al. [11], and the other is the image generated by the proposed method. Because the front region near the operational part is occluded by the operational part itself, it is impossible to present for the image generated by Sato's method, which uses outward-facing fisheye cameras only, to visualize the front region. In contrast, the image given by proposed method shows the front region of the operational part. This area is the main focus of the operator when moving the operational part. For quantitative assessment, each corner point was picked up manually, then the angle of corner (θ_i) and the length of edge of each tetragon (l_i) was computed. The number of visible vertex ($n_{visible}$) of tetragons was also counted.

Ideally, all of the corner is right angled and all of edge has equal length, when the tetragon is observed from above with a perspective camera. Thus, identically, following condition must be satisfied:

$$l_i = l_j \quad (10)$$

$$\theta_i = 90 \text{ deg for all } i, j \quad (11)$$

The larger number of visible vertex of tetragons means the wider ground region can be observed. Figure 11 was evaluated by the number of visible corner $n_{visible}$, relative error of the length of the edges of each tetragon E_l , the average of angle of corners $\bar{\theta}$, and the standard deviation of angle σ_θ . The definitions of these indexes are the following



(a) Image acquired by the field camera

(b) Image generated by the method of paper [11]

(c) Image generated by the proposed method

Fig. 11.

equations:

$$\bar{l} = \frac{1}{N} \sum l_i \quad (12)$$

$$\sigma_l = \sqrt{\frac{1}{N} \sum (l_i - \bar{l})^2} \quad (13)$$

$$E_l = \frac{\sigma_l}{\bar{l}} \quad (14)$$

$$\bar{\theta} = \frac{1}{N} \sum \theta_i \quad (15)$$

$$\sigma_\theta = \sqrt{\frac{1}{N} \sum (\theta_i - \bar{\theta})^2} \quad (16)$$

$$(17)$$

If ideal image is evaluated, E_l becomes 0, $\bar{\theta}$ is equal to 90degree, and σ_θ becomes 0.

Table I shows the result of evaluation on Fig. 11.

This table shows that the image generated by the proposed method is superior to the image generated by the conventional method, because the length error and the standard deviation of corner angle of the former is smaller than these of the latter, and the number of visible vertex of the former is larger than that of the latter.

C. Teleoperation experimental results

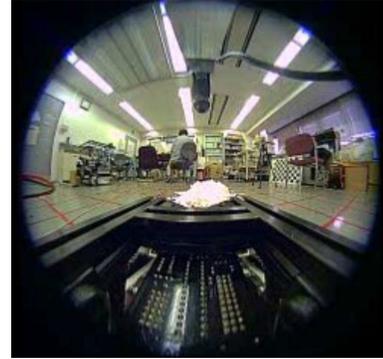
In the experiment, the bulldozer was controlled to push pieces of Styrofoam (the work object), which we liken to earth and soil. Figure 12 shows the experimental setting.

The presented image given by the proposed method was compared to actual images acquired by cameras. Figure 12(b) shows images acquired by front fisheye camera. It is impossible to see the situation of the operational part, which is essential for smooth teleoperation. For example, in the situation shown in Fig. 12(a.), none of the Styrofoam is in the operation part. However, the difference of amount of work object in operational part cannot be seen in Fig. 12(b).

Figure 13 shows the image presented to the operator by the proposed method. In this figure, it is clearly possible to see the surrounding environment and the operational part.



(a) Field camera image



(b) Front fisheye camera

Fig. 12. Experimental images

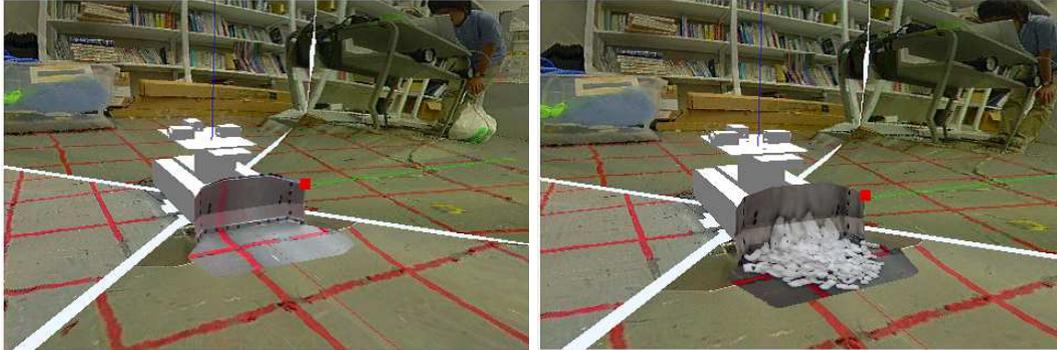
D. Test on real machine

The proposed system was implemented to a bulldozer(Komatsu D155A) to make sure that the system works on the real situation. Figure 14 shows how the cameras were placed.

Four fisheye cameras which acquire the surrounding environment were placed on the front, back, left, and right of the bulldozer, and a perspective camera was placed on the tip of the pole.

TABLE I
GEOMETRICAL EVALUATION RESULT

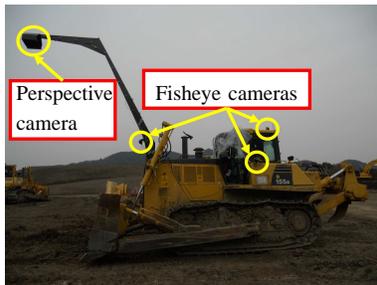
	Field camera	Conventional method	Proposed method
Length error (E_l)	3.0×10^{-2}	18.7×10^{-2}	15.9×10^{-2}
Average angle (θ)[deg]	90.0	90.5	89.7
Standard deviation of corner angle (σ_θ)[deg]	2.9	17.5	14.7
Number of visible vertex ($n_{visible}$)	383	271	283



(a) operational part only

(b) operational part and work objects

Fig. 13. Images presented by the proposed method



(a) Overview



(b) Front fisheye camera



(c) Left fisheye camera



(d) Back fisheye camera



(e) Right fisheye camera



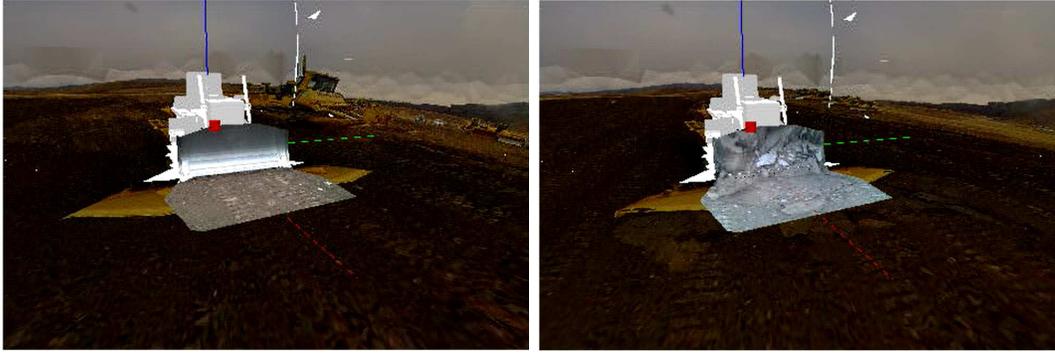
(f) Perspective camera acquiring operation part

Fig. 14. The location of cameras on the real machine

Figure. 15 shows the images presented to the operator by the proposed method.

IV. CONCLUSION

We have proposed a new image presentation method that incorporates the surrounding environment and the operational



(a) operational part only

(b) operational part and work objects

Fig. 15. Images presented by the proposed method

part. The proposed method is based on the texture mapping to a sphere, and uses a 3DCG model of the operational part and work object. In particular, projection matrix for the mapping from the camera observing the operational part to 3DCG model is obtained without the needs for any additional objects. The proposed method can visualize the region in front of the operational part, in addition to the surrounding environment. This front region is not adequately represented by previous methods [11]. In future work, to verify the effectiveness of the proposed method for teleoperation, by evaluating the speed and accuracy of teleoperation.

REFERENCES

- [1] S. Kawatsuma, M. Fukushima, and T. Okada, "Emergency response by robots to fukushima daiichi accident: summary and lessons learned," *Industrial Robot: An International Journal*, vol. 39, no. 5, pp. 428–435, 2012.
- [2] Y. Segawa, Y. Nanamori, S. Sanji, T. Kitada, and S. Tadokoro, "An introduction to japanese r&d activity and political direction on special environment robots," in *Humanitarian Technology Conference (R10-HTC), 2013 IEEE Region 10*, 2013, pp. 29–34.
- [3] J. Chen, E. Haas, and M. Barnes, "Human performance issues and user interface design for teleoperated robots," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 37, no. 6, pp. 1231–1245, Nov 2007.
- [4] A. Valtzanos and S. Ramamoorthy, "Evaluating the effects of limited perception on interactive decisions in mixed robotic domains," in *Proceedings of the 8th ACM/IEEE International Conference on Human-robot Interaction*. Piscataway, NJ, USA: IEEE Press, 2013, pp. 9–16.
- [5] C. James, T. Bednarz, K. Hausteine, L. Alem, C. Caris, and A. Castleden, "Tele-operation of a mobile mining robot using a panoramic display: an exploration of operators sense of presence," in *Automation Science and Engineering (CASE), 2011 IEEE Conference on*, Aug 2011, pp. 279–284.
- [6] T. Sanguino, J. Marquez, T. Carlson, and J. Del R Millan, "Interaction and evaluation of an augmented virtuality assistance system for teleoperated robots," in *Robotic and Sensors Environments (ROSE), 2012 IEEE International Symposium on*, Nov 2012, pp. 19–24.
- [7] C. Nielsen, M. Goodrich, and R. Ricks, "Ecological interfaces for improving mobile robot teleoperation," *Robotics, IEEE Transactions on*, vol. 23, no. 5, pp. 927–941, Oct 2007.
- [8] S. Hashimoto, A. Ishida, M. Inami, and T. Igarashi, "Touchme: An augmented reality based remote robot manipulation," in *The 21st International Conference on Artificial Reality and Telexistence, Proceedings of ICAT2011*, 2011.
- [9] F. Okura, Y. Ueda, T. Sato, and N. Yokoya, "Teleoperation of mobile robots by generating augmented free-viewpoint images," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, Nov 2013, pp. 665–671.
- [10] Y.-C. Liu, K.-Y. Lin, and Y.-S. Chen, "Bird's-eye view vision system for vehicle surrounding monitoring," in *Robot Vision*, ser. Lecture Notes in Computer Science, G. Sommer and R. Klette, Eds. Springer Berlin Heidelberg, 2008, vol. 4931, pp. 207–218.
- [11] 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," in *System Integration (SII), 2013 IEEE/SICE International Symposium on*, 2013, pp. 753–758.
- [12] J. Tran, A. Ufkes, M. Fiala, and A. Ferworn, "Low-cost 3d scene reconstruction for response robots in real-time," in *Safety, Security, and Rescue Robotics (SSRR), 2011 IEEE International Symposium on*, Nov 2011, pp. 161–166.
- [13] A. Kelly, N. Chan, H. Herman, D. Huber, R. Meyers, P. Rander, R. Warner, J. Ziglar, and E. Capstick, "Real-time photorealistic virtualized reality interface for remote mobile robot control," *The International Journal of Robotics Research*, vol. 30, no. 3, pp. 384–404, 2011.