A Tele-Operated Display With a Predictive Display Algorithm

YEONGTAE JUNG, KYUTAEK HAN, AND JOONBUM BAE, (Member, IEEE)
Department of Mechanical Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, South Korea

Corresponding author: Joonbum Bae (jbbae@unist.ac.kr)

ABSTRACT Tele-operated display systems with head mounted displays (HMD) are becoming popular as visual feedback systems for tele-operation systems. However, the users are suffered from time-varying bidirectional delays caused by the latency and limited bandwidth of wireless communication networks. Here, we develop a tele-operated display system and a predictive display algorithm allowing comfortable use of HMDs by operators of tele-operation systems. Inspired by the kinematic model of the human head-neck complex, we built a robot neck-camera system to capture the field of view in any desired orientation. To reduce the negative effects of the time-varying bidirectional communication delay and operation delay of the robot neck, we developed a predictive display algorithm based on a kinematic model of the human/robot neck-camera system, and a geometrical model of a camera. Experimental results showed that the system provide predicted images with high frame rate to the user.

INDEX TERMS Tele-operateion, head mounted display, delay compensation.

I. INTRODUCTION

Sites of disasters such as natural disasters, radioactive accidents and chemical accidents, which can pose problems in terms of human access, are increasing [2], [3]. Tele-operation systems have been researched to work in such disaster sites with human user's intelligence [4]–[7]. In the tele-operated systems, the user typically obtains the information of the sites from two-dimensional images transmitted to monitors by cameras attached to the tele-operated robot. Visual data is the most intuitive form of information when observing environments. However, the performance of a tele-operated system is limited if three-dimensional (3D) stereoscopic images are not provided, because the user cannot gauge distances between objects.

Recently, head mounted displays (HMDs) affording immersive 3D visual feedback have been used to observe environments and control tele-operated robots to perform manipulations [8]–[10]. Most HMDs feature integrated inertial measurement units (IMUs) that measure the orientation of the user's head. The head orientation is used to capture the field of view (FOV) in the direction in which the user looks by

The associate editor coordinating the review of this manuscript and approving it for publication was Zheng Chen.

employing a robot neck-camera system or a panoramic camera of tele-operated robots. However, it is difficult to provide real-time images in such tele-operated situations; the latency, limited bandwidth of the tele-communication network, and the large sizes of stereoscopic images impose delays and packet loss. Such delays and loss of data induced by unstable nature of wireless tele-communication network trigger time-varying delays in image presentation, causing simulator sickness [11]–[13]. The operation delays imposed by the physical limitations of robot neck-camera system increase the delay further.

Delays in tele-operation system have been researched for decades, especially to ensure stability and transparency of bilateral haptic systems. Many control algorithms, such as wave-variable based passivity control algorithms [14], [15] and modified algorithms to address constant [16], [17] and time-varying delays [18]–[20] have been researched. Adaptive and robust control algorithms have been studied to deal time-delayed systems with nonlinearities and uncertainties [21]–[24]. Despite of the achievements, it is difficult to apply such methods in visual feedback systems because of different nature.

The use of point cloud data or images to construct virtual worlds (rather than images recently captured by the robot

camera) may mitigate the delay effect problems [9], [25], [26]. However, such methods impose large computational burdens or require expensive sensors. The delay effect could be reduced if the user's head motion is predicted, and the remotely controlled robot moves ahead of the user. Prediction of user's head motion have been researched to reduce the delay of HMDs for displaying virtual reality [27], [28]. However, it is difficult to apply such predictions, because large delays and packet loss are expected in the communications under disaster situations. Model-based predictions have been researched to mitigate the effect of delays in tele-operation systems. The position and orientation of robot manipulators were predicted by kinematic models of the robots and reflected in the user-side display, to avoid long task time induced by ''move and wait'' strategy [29], [30] or increase manipulation precision [31]. However, user's head motion was not involved in the systems, because the display systems consisted of ordinary monitors. A predictive display method was suggested in [32] to provide the image in actual direction of a tele-operated vehicle utilizing the dynamics of the vehicle, but it was not implemented in real system.

Here, we developed a tele-operated display system and a practical predictive display algorithm that compensates for the bidirectional network and operation delays. We constructed a robot neck-camera system based on a kinematic model of the human head-neck complex. The display algorithm predicts images in the direction of user's head orientation by employing delayed image and analyzing the difference between the delayed robot neck orientation and the current user head orientation. We compensated for the bidirectional communication and operational delays by translating and rotating the delayed images using kinematic models of the human neck and the robot neck-camera system, and a geometrical model of the camera. The delays were addressed by predicting images which correspond to the current user's head orientation, utilizing the delayed images, robot neck orientation, and current user's head orientation.

The remainder of this paper is organized as follows. Section II provides an overview of our tele-operated display system. Section III contains a detailed analysis of the camera and the predictive display algorithm. The experimental setup and the results afforded by the proposed algorithm are shown in Section IV. Section V presents the conclusions and describes planned future work.

II. OVERVIEW OF THE DISPLAY SYSTEM

We developed an intuitive display system for tele-operated robots. As shown in Figure 1, the human head-neck complex was modeled as a simplified series linkage system featuring three revolute joints to represent the rotational head motion. The HMD [33] attached to the user's head measures head orientation, allowing presentation of 3D stereoscopic images. The robot neck had the same kinematic structure as the human model. Two cameras [34] were placed on top of the robot neck to capture stereoscopic view of the work site. To prevent user's sickness caused by time differences between the


▲A kinematic model of the human head-neck complex

▲ A robot neck-camera module for the 3D vision feedback.
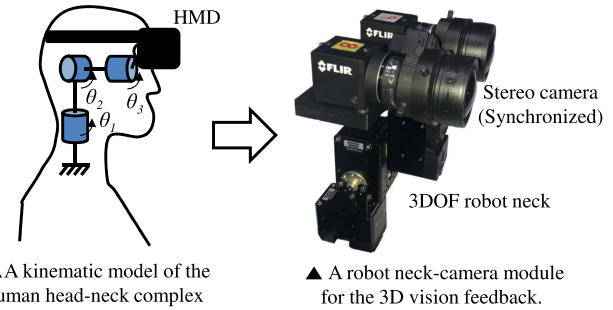
FIGURE 1. Design of a robot neck-camera module based on the kinematic model of human head-neck complex. (θ_1, θ_2 and θ_3 are the rotation angle of each rotational joints that represents yaw, pitch and roll motion of the head, respectively).

TABLE 1. Specifications of the tele-operated display system.

	Specification
HMD	Oculus VR, Oculus Rift
	Display Resolution : 960 x 1080 per eye
	IMU : InvenSense, MPU-9250
Motor	ROBOTIS, XH430-V210-R
	Gear ratio : 212.6:1
	No load speed : 52 RPM
	Stall Torque : 2.6 Nm
	Operation Range : -180° to 180°
Encoder	ams, Contactless absolute encoder
	Resolution : 12 bit
Camera	FLIR, FL3-U3-32S2C-CS
	Image sensor : Sony IMX036
	Maximum resolution : 2080 x 1552
	Maximum frame rate : 60 FPS
	Lens : FUJIFILM, YV2.8X2.8SA-2
Computer	Processor : Intel, Core i7-4770
(User side)	Graphic Driver : NVIDIA, GTX 1060
	RAM : 8GB

stereoscopic cameras, the cameras were synchronized by digital signals; both captured images at the same moment. The detailed specifications of the tele-operated display system are listed in Table 1.

Figure 2 is a schematic of the system. The orientation of the user's head ($o_H(t)$) is measured by an IMU inside the HMD, with a measurement delay of Δ_{IMU}. The measured orientation [$o_H(t - \Delta_{IMU})$] is delivered to the robot neck-camera through a wireless network with a communication delay of Δ_{CM}, and the robot neck then follows the user's head orientation [$o_H(t - \Delta_{IMU} - \Delta_{CM})$] with an operational delay Δ_{OP} caused by physical limitations of the actuators. The cameras captures stereo images in the direction of their orientation $o_R(t)$, which is the same as the human neck orientation but delayed by Δ_{IMU}, Δ_{CM} and Δ_{OP}; the delayed orientation is denoted as $o_H(t - \Delta_{IMU} - \Delta_{CM} - \Delta_{OP})$. The captured images [Image($o_R(t)$)] and the orientations of image capture [$o_R(t)$] are delivered to the user with an additional communication delay of Δ_{CM}.
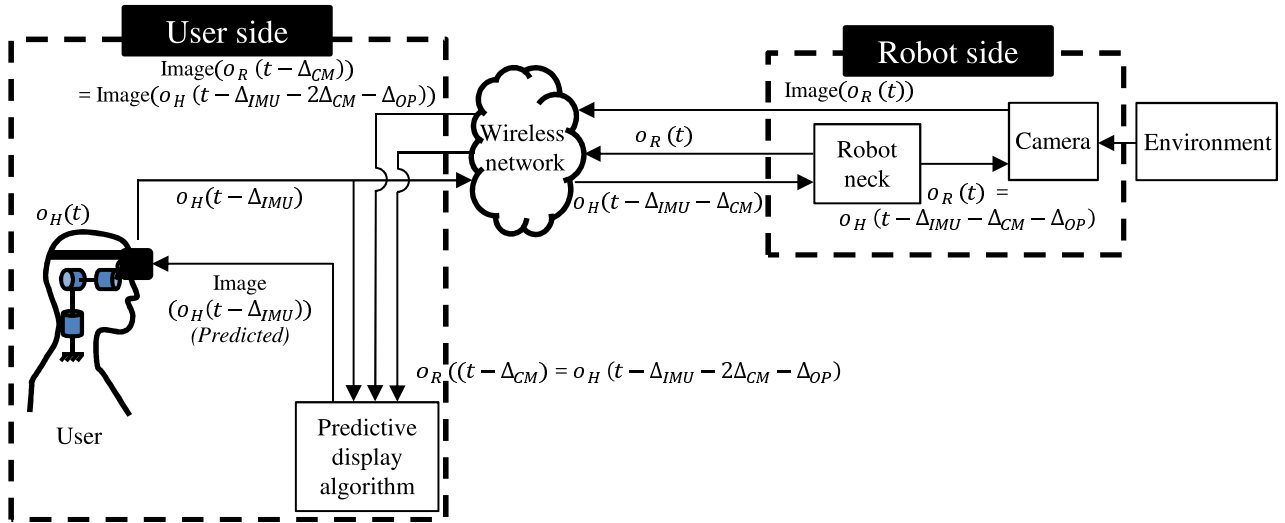
FIGURE 2. Overview of the tele-operated display system. o_H and o_R are the orientation of user's head and robot neck-camera module, respectively. Δ_{IMU}, Δ_{CM} and Δ_{OP} represent the measurement delay of user's head orientation, communication delay of the wireless network and operational delay of the robot neck, respectively.

The images can be delivered directly to the user via the HMD. However, as shown in Figure 2, the delivered images [Image($o_R(t - \Delta_{CM})$) = Image($o_H(t - \Delta_{IMU} - 2\Delta_{CM} - \Delta_{OP})$)] contain the measurement delays of the IMU [Δ_{IMU}], bidirectional communication delay [$2\Delta_{CM}$], and the operational delay of the robot neck [Δ_{OP}]. If the image is provided directly to the user, the user may suffer because of the delays. Especially, bidirectional communication delay is time-varying in nature and can trigger simulator sickness [11]–[13]. Also, random loss of image data, which is frequent because of the size of stereoscopic images, increases variation in Δ_{CM}. Such delays and image loss are inevitable in wireless networks, unless the communication is exceptionally well-controlled. Δ_{OP} is also time-varing, as it depends on user's head motion. Δ_{CM} and Δ_{OP} are the predominant delays; Δ_{IMU} is both relatively small and constant. To reduce the undesirable effects of Δ_{CM} and Δ_{OP} (such as sickness [11]–[13]), we develop a predictive compensation algorithm. The algorithm modifies the delivered images [Image($o_R(t - \Delta_{CM})$)] using delayed robot neck orientations [$o_R(t - \Delta_{CM})$] and the measured user head orientation [$o_H(t - \Delta_{IMU})$] before providing the images. In this way, the image which is supposed to be in the direction of user's head is predicted; thus, the provided image instantly reflects user's head motion despite the existence of time-varying delays. The algorithm compensates for Δ_{CM} and Δ_{OP}, which are the dominant time-varying delays of the entire system that affect to the sickness of users.

III. THE PREDICTIVE DISPLAY ALGORITHM
A. DERIVATION
We first analyzed camera geometry. In this analysis, we assumed that the camera is a pinhole camera without lens distortion. Also, the image sensor was assumed to be square.

The camera was analyzed by dividing robot neck motions into yaw, pitch motion (θ_1 or θ_2 in Figure 1, respectively) and roll motion (θ_3 in Figure 1) of the robot neck. Figure 3 shows the simplified geometry; the camera features an aperture and an image sensor. Figure 3a shows image formation on the image sensor and the position change of the formed image by yaw or pitch motion of the robot neck. If a subject is placed in the direction of θ from the perpendicular line of the sensor, the distance from the center of the image sensor to the formed subject image [$\overline{op}$] can be approximated as follows:

$$\overline{op} = dtan\theta \approx d\theta \tag{1}$$

where d is the distance between the aperture and the camera image sensor. If the camera is rotated by θ', the distance changes to $\overline{op'}$, as follows:

$$\overline{op'} = dtan(\theta + \theta') \approx d(\theta + \theta') \tag{2}$$

Using (1) and (2), the positional change in a formed image [$\overline{pp'}$] can be calculated as:

$$\overline{pp'} = \overline{op'} - \overline{op} = d\theta' \tag{3}$$

Note that $\overline{pp'}$ is proportional to the translation distance of the subject in the captured image (in pixels, δ_{pixel}) as follows:

$$\delta_{pixel} = \alpha d\theta' \tag{4}$$

where α is a conversion factor used to transform the distance change of the subject image on the image sensor to that in the captured image in pixels. Figure 3b shows image formation of a subject on the image sensor and its position change due to the roll motion by θ'. In such a case, the image sensor also rotates by θ', as does the subject in the captured image.

The relationship between changes in camera orientation and positional change of the subject in captured images can be used to predict future images based on current images.

(a) Image position change by the yaw or pitch motion.

(b) Image position change by the roll motion.
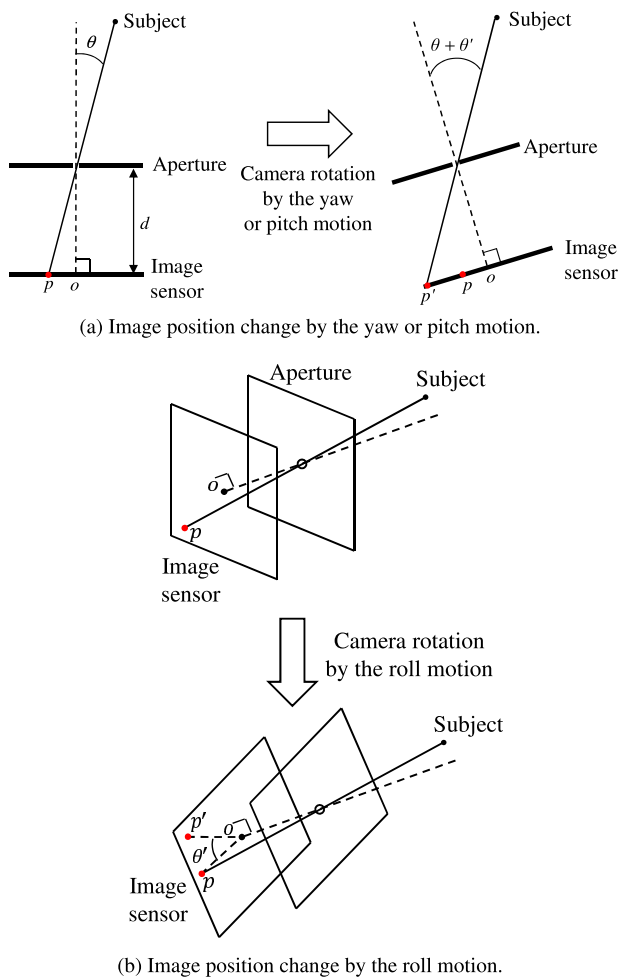
FIGURE 3. Simplified camera geometry and image position change of a subject by camera motions.

If the camera joints corresponding to θ_1, θ_2 and θ_3 of the human kinematic model are rotated by θ'_1, θ'_2 and θ'_3 respectively, the future image can be predicted by translating the current image by $\alpha d\theta'_1$ in the horizontal direction and $\alpha d\theta'_2$ in the vertical direction, and by rotating the image through θ'_3. Similarly, the scene in the direction of the current user head orientation can be predicted using that current orientation of user's head [$o_H(t - \triangle_{IMU})$], the camera orientation when the delayed image was captured [$o_R(t - \triangle_{CM})$], and the delayed image [Image($o_R(t - \triangle_{CM})$)].

The image prediction can be implemented by introducing a cropped area, and translating or rotating it by the orientation difference between the robot and human for yaw, pitch and roll motion, respectively. An example of image manipulation in the case of yaw motion (rotation of θ_1) is shown in Figure 4a. Here, a single axis-rotation case is considered, for the clarity of the explanation. The camera captures an image in the direction of its current orientation, as measured by the robot neck system, and both the image [Image($o_R(t)$)] and the robot neck orientation [$o_R(t)$] are delivered to the user side with a delay of $\triangle_{CM}$. If the predictive display algorithm

is not applied, the delivered image does not change until a new image arrives. Thus, the delay generates an orientation difference between the user head orientation [$o_H(t - \triangle_{IMU})$] and the currently displayed image's camera orientation [$o_R(t - \triangle_{CM}) = o_H(t - \triangle_{IMU} - 2\triangle_{CM} - \triangle_{OP})$].
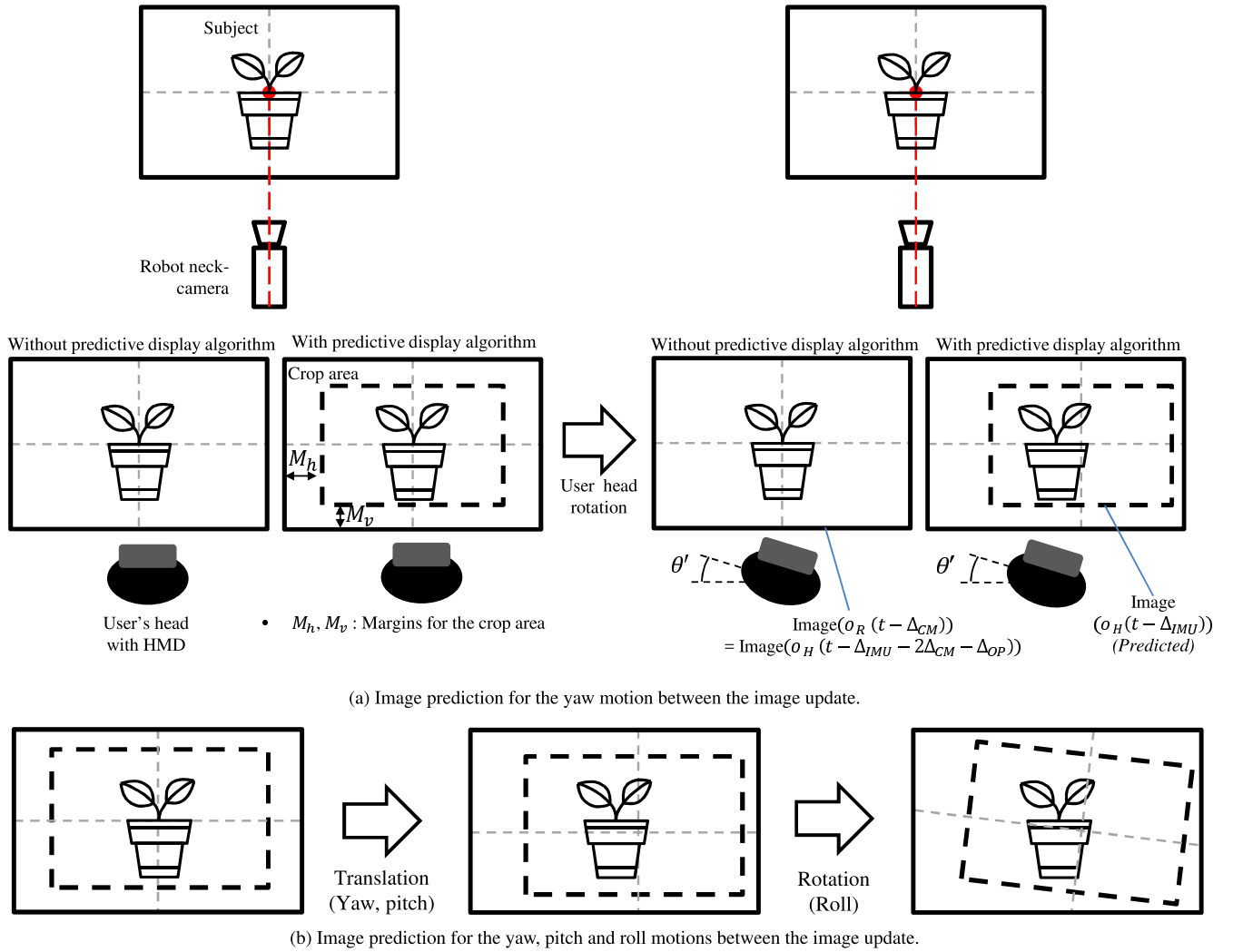
However, the predictive display algorithm provides a manipulated image (not an original) to the user. A rectangular cropping area imposed on the delivered image [Image($o_R(t - \triangle_{CM})$)] with horizontal and vertical margins of M_h and M_v, respectively. If the user head orientation changes, the cropped area is translated by $\alpha d\theta'$ pixels in the direction of head rotation, where θ' is the difference between the current user head orientation [$o_H(t - \triangle_{IMU})$] and the delayed robot neck orientation [$o_R(t - \triangle_{CM})$]. The translated cropped image is then provided to the user through the HMD; the extent of translation is determined by the camera model of (4).

Similarly, the algorithm can deal with pitch and roll motion of the user head by applying the manipulation sequentially (Figure 4b). The cropped areas are translated horizontally and vertically by $\alpha d\theta'_1$ and $\alpha d\theta'_2$, respectively, and rotated by θ'_3, where θ'_1, θ'_2 and θ'_3 are the yaw, pitch and roll components of the difference between $o_H(t - \triangle_{IMU})$ and $o_R(t - \triangle_{CM})$. As the algorithm predicts images in the direction of current user head orientation [$o_H(t - \triangle_{IMU})$], the effect of $\triangle_{CM}$ and $\triangle_{OP}$ is compensated. Thus, image immediately change following user head rotation feature a delay of only $\triangle_{IMU}$, which is both constant and much smaller than the time-varying delays $\triangle_{CM}$ and $\triangle_{OP}$. Therefore, the predictive display algorithm can reduce the discomfort feelings or the sickness of the display system caused by delays.

A delay compensation algorithm was proposed in Edwards' patent with a similar approach [35]. However, that method yields an image smaller than or equal to the FOV of the HMD, with negative or zero M_h and M_v values. Thus, the image shown in the HMD looks like a scene through a window that is translated and rotated when the image is updated. This may negatively affect user immersion, limiting the user-side FOV. Although the concept is similar to ours, there was no analysis about the camera geometry and image position change; our analyses support the validity of our algorithm. Also, the extent of margins were not considered; we discuss these margins in Section III.B.

B. MARGIN ANALYSIS
The margins must be set considering the rotational speed of the user's head and the image delays caused by $\triangle_{CM}$ and $\triangle_{OP}$. If the cropping area is overlapped with the outside of image, the user no longer receives a square image, but rather a clipped image, because information is lacking outside the image. This compromises task performance of the tele-operation system by reducing user immersion. Thus, the margins M_h and M_v must be large enough to make the cropped area do not attain the image edges. However, large margins reduce the image FOV in turn reducing the information imparted. Given this trade-off, the margins must be as

(a) Image prediction for the yaw motion between the image update.

(b) Image prediction for the yaw, pitch and roll motions between the image update.

FIGURE 4. Implementation procedure of the predictive display algorithm.

small as possible but sufficiently large to prevent overlapping with cropped areas the outside the image.

The required margins can be calculated from the image delay, and the rotational speeds of the user's head in yaw ($\dot{\theta}_1$), pitch ($\dot{\theta}_2$) and roll ($\dot{\theta}_3$). Figure 5 shows the translation and rotation of the cropped area. The four corners are denoted A, B, C, and D. As one of these points will be the first point to reach the edge of the image when the cropped area begins to overlap with and area outside the image, it is necessary to ensure that all points stay within the image. Margin analysis commences at point A with initial coordinates (x_A, y_A). The distances between A and closest horizontal and vertical image edges are denoted $M_{h,A}$ and $M_{v,A}$, respectively. The coordinates of point A prior to translation and rotation of the cropped area are:

$$x_A = l cos\theta_i \qquad (5)$$
$$y_A = l sin\theta_i \qquad (6)$$

where l is the distance from the center of the image to A, and θ_i is the angle between the horizontal line and $\overline{OA}$, calculated

using the margins and image size, as follows:

$$l = \sqrt{(L_h/2 - M_h)^2 + (L_v/2 - M_v)^2} \qquad (7)$$
$$\theta_i = atan\left(\frac{L_v/2 - M_v}{L_h/2 - M_h}\right) \qquad (8)$$

Before manipulation of the cropped area, $M_{h,A}$ and $M_{v,A}$ are identical to M_h and M_v. As the area is translated and rotated, the coordinates of point A change:

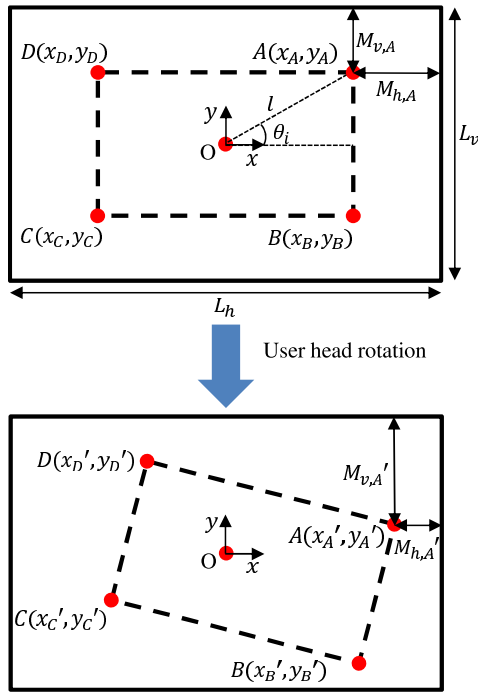$$x'_A = l cos(\theta_i + \theta'_3) + \alpha d\theta'_1 \qquad (9)$$
$$y'_A = l sin(\theta_i + \theta'_3) + \alpha d\theta'_2 \qquad (10)$$

Assuming an image delay of dt and user rotational speeds of $\dot{\theta}_1, \dot{\theta}_2, \dot{\theta}_3$ (yaw, pitch and roll axis, respectively), the positional change of point A caused by user motion during the delay time dt can be calculated as follows:

$$x'_A = l cos(\theta_i - \dot{\theta}_3 dt) - \alpha d\dot{\theta}_1 dt \qquad (11)$$
$$y'_A = l sin(\theta_i - \dot{\theta}_3 dt) + \alpha d\dot{\theta}_2 dt \qquad (12)$$

FIGURE 5. Margin changes by the translation and rotation of the cropped area.

FIGURE 6. Experimental setup for the parameter identification.

such that the cropped area does not reach to the edge of the image. Using these criteria, the minimal required margins can be calculated by applying the maximum angular velocity of the user's head ($\dot{\theta}_{1,max}$, $\dot{\theta}_{2,max}$ and $\dot{\theta}_{3,max}$) and maximum image delay (dt_{max}) into (11)-(12) and (16)-(21). On the contrary, the delay that can be dealt with a set of margins can be obtained by assigning the margins to (7)-(8) and calculating dt that satisfies (15) and (28).

The problem here is that the image delay changes significantly as tele-communication network conditions vary. To address this problem, margins can be calculated and applied in realtime, because our predictive display algorithm does not impose a large computational burden. However, margin variation also changes the image FOVs, which may disturb the user. Also, there are limitations in the margins, since the image have limited size. If the user's head motion is fast enough to escape the margins in the given network delay, reaching to the outside of image is unavoidable. Thus, we determined the margins empirically in the following performance tests. The margins were first set to zero, and then were increased until the cropped area does not reach to the edge of image during the experiments so that the margin criteria can be satisfied under given delay and user's rotational speed in a test operation.

The positional change varies the margins $M_{h,A}$ and $M_{v,A}$ as follows:

$$M'_{h,A} = L_h/2 - x'_A \quad (13)$$
$$M'_{v,A} = L_v/2 - y'_A \quad (14)$$

To ensure that the cropped area does not attain the edges of the image, $M_{h,A}$ and $M_{v,A}$ must satisfy the following criteria:

$$M'_{h,A}, M'_{v,A} > 0 \quad (15)$$

Similarly, the coordinates and margins of the other three points after manipulation of the cropped area are:

$$x'_B = lcos(\theta_i + dot\theta_3 dt) - \alpha d dot\theta_1 dt \quad (16)$$
$$y'_B = -lsin(\theta_i + dot\theta_3 dt) + \alpha d dot\theta_2 dt \quad (17)$$
$$x'_C = -lcos(\theta_i - dot\theta_3 dt) - \alpha d dot\theta_1 dt \quad (18)$$
$$y'_C = -lsin(\theta_i - dot\theta_3 dt) + \alpha d dot\theta_2 dt \quad (19)$$
$$x'_D = -lcos(\theta_i + dot\theta_3 dt) - \alpha d dot\theta_1 dt \quad (20)$$
$$y'_D = lsin(\theta_i + dot\theta_3 dt) + \alpha d dot\theta_2 dt \quad (21)$$
$$M'_{h,B} = L_h/2 - x'_B \quad (22)$$
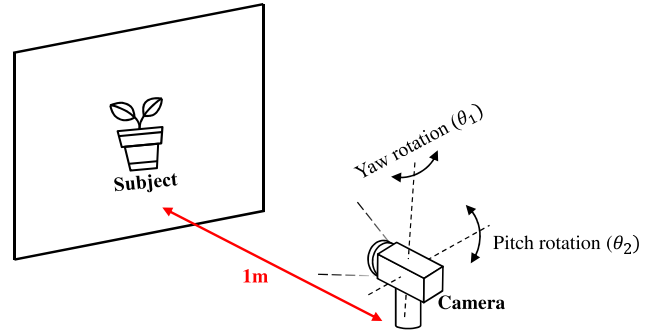$$M'_{v,B} = L_v/2 - y'_B \quad (23)$$
$$M'_{h,C} = L_h/2 - x'_C \quad (24)$$
$$M'_{v,C} = L_v/2 - y'_C \quad (25)$$
$$M'_{h,D} = L_h/2 - x'_D \quad (26)$$
$$M'_{v,D} = L_v/2 - y'_D \quad (27)$$

$M_{h,B}$, $M_{v,B}$, $M_{h,C}$, $M_{v,C}$, $M_{h,D}$ and $M_{v,D}$ are also required to satisfy following criteria:

$$M'_{h,B}, M'_{v,B}, M'_{h,C}, M'_{v,C}, M'_{h,D}, M'_{v,D} > 0 \quad (28)$$

IV. EXPERIMENTS
A. PARAMETER IDENTIFICATION
To apply the predictive display algorithm, the camera parameter αd of (4) was identified by an experiment. Figure 6 shows the experimental setup. The robot neck-camera module was fixed, and a subject was placed in front of the camera at a distance of 1 m, occupying the center of the image when the neck of the robot was in the initial position. The robot neck was rotated (in yaw or pitch) from $-30°$ to $30°$ in $1°$ steps with the camera operating. The distance between the subject in the captured image and the image center was recorded in pixels; the αd parameter was identified by linear fitting of αd to (4), applying the least squares method to the recorded position and rotational angle. Thus, αd was identified as $\alpha d = 5.5$ pixel/deg for yaw, and $\alpha d = 5.9$ pixel/deg for pitch with root mean square (RMS) errors of 3.3 pixel and 4.1 pixel respectively. Since there was no significant outlier, the identified model showed good agreement with the measured data with

(a) Yaw motion.
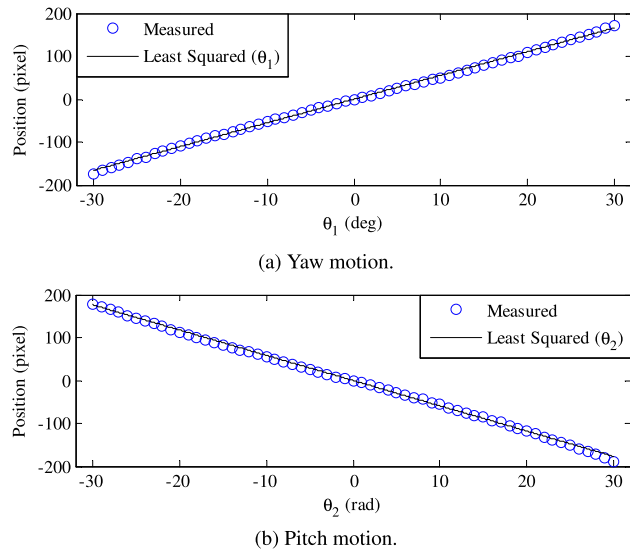
(b) Pitch motion.

FIGURE 7. **Identification result of camera parameter** αd **by yaw and pitch motions.**
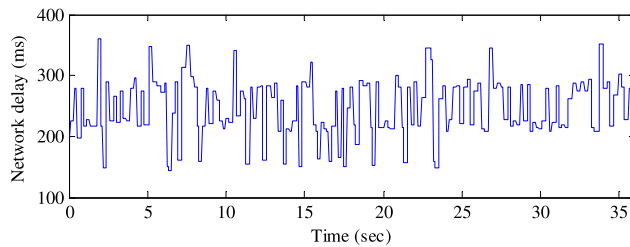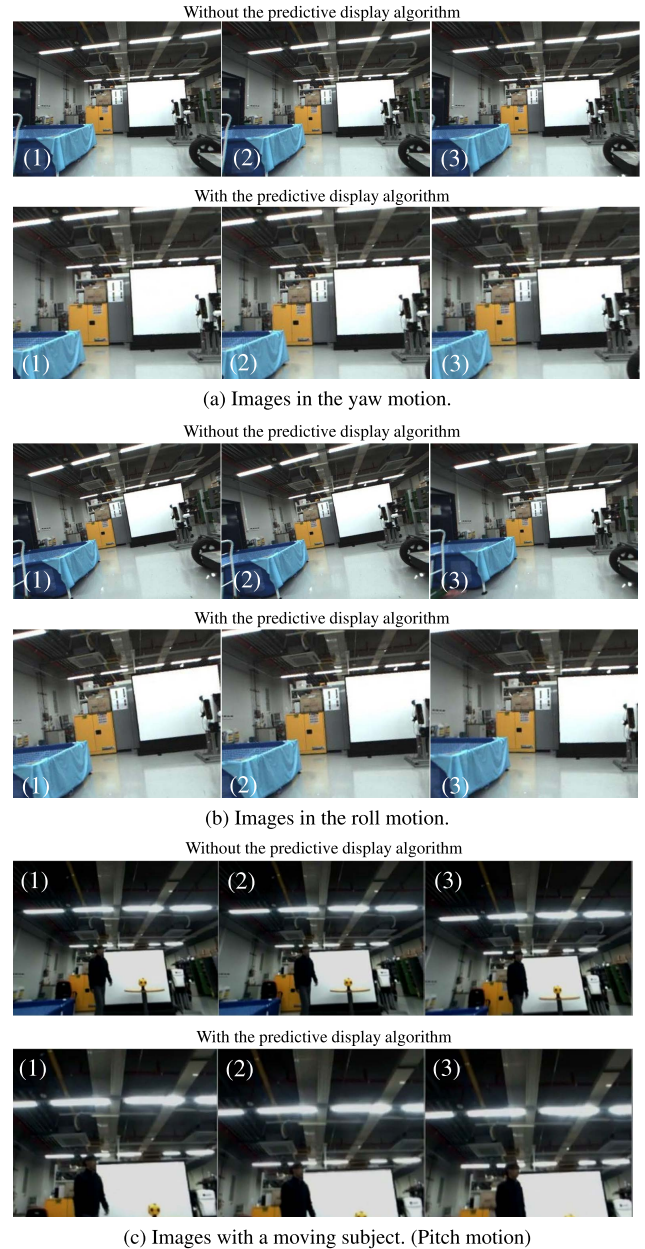

FIGURE 8. **A random network delay.**

the least squares method. Figure 7a and Figure 7b show the measured positions of the subjects and the linearly fitted models for yaw and pitch of the robot neck, respectively. As the graphs show, the measured data and camera model of (4) were in good agreement. The parameters differed slightly in terms of both yaw and pitch motion, caused by lens distortion or/and image sensor asymmetry. The parameters depend on the physical specifications of the lens and the image sensor of camera, such as the refractive index of the lens and the size and position of the image sensor, which are not time-varying specifications. Thus, pre-identified camera parameters were used in the implementation of predictive display algorithm.

B. IMPLEMENTATION OF THE PREDICTIVE DISPLAY ALGORITHM

The predictive display algorithm was tested in a real-world tele-operation experiment. The experimental setup was the same as that of Figure 2. The user wore the HMD and rotated the head; head orientation was measured and delivered to the robot neck-camera module through a wireless network; the robot neck followed the user's motion while a stereo camera captured environmental images at a resolution of 554×413 pixels. The captured images and robot neck

(a) Images in the yaw motion.

(b) Images in the roll motion.

(c) Images with a moving subject. (Pitch motion)

FIGURE 9. **Captured images in the tele-operation experiment. Focus on changes of image edges (e.g., the robot on the right side), or the moving subject (i.e., the human). Check the attached video file for more details.**

orientations when the images were taken were delivered to the user side. A random delay was intentionally added to the wireless network to reflect network conditions in practical tele-operated applications. The captured image and robot neck orientations were sent to the user side after random delay with a maximum magnitude of 230 ms and a minimum of 70 ms, representing a communication delay in the tele-operated display system. Both images (i.e., those subjected to and not subjected to predictive display algorithm) were recorded. The margins of the predictive display algorithm were set to appropriately tuned values ($M_h = 90$, $M_v = 50$) as in Section III, or inappropriate values ($M_h = 0$, $M_v = 0$
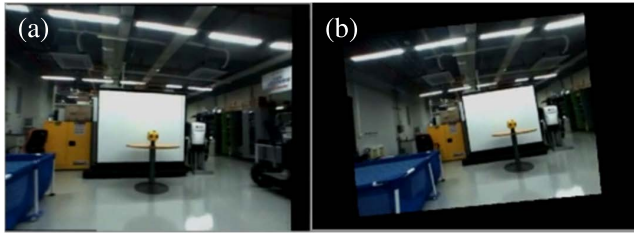
FIGURE 10. Experimental results with inappropriate margins. ((a): Zero margins, (b): Negative margins) Check the attached video file for more details.

(a) Feature matching of the captured images. (With the predictive display algorithm)

(b) Average keypoint coordinate differences between images before and after update.

FIGURE 11. Feature analysis results.

and $M_h = -90$, $M_v = -50$) as suggested in Edwards' patent [35] Given that the frame rate of most digital videos is 30 frames per second (FPS), image prediction proceeded at in 30 Hz.

The recorded results are attached as a supplementary video file. The bidirectional network delay is shown in Figure 8. The minimum, average, maximum, and standard deviation of the delay were 145, 245, 361, and 48 ms, respectively. Figure 9 shows sets of updated images from scenes (1) to (3) without/with the predictive algorithm operating under the same experimental conditions with appropriate margins. In the absence of the predictive display algorithm, the captured images do not change until new images arrive. However, the images modified by the predictive algorithm change continuously as the user's head rotates. As shown in the video file, the delay-compensated images are much smoother and more natural than those presented in the absence of the algorithm. The desired frame rate (30 FPS) was maintained even in the presence of unpredictable time-varying delays. The results imply that the proposed algorithm allows operators of tele-operation systems to comfortably use their HMDs.

A feature analysis was performed to verify the performance of proposed algorithm. In this analysis, the coordinate differences of keypoints between the images before and after the arrival of new image were compared, as shown in Fig. 11a. A fast library for approximate nearest neighbors (FLANN) matcher in openCV was used in the analysis. The average values of coordinate differences are plotted in Fig. 11b. The average difference was 11.0 pixels and 1.9 pixels for the images without the predictive display algorithm and with the algorithm, respectively. The coordinate difference was significantly reduced when the predictive display algorithm was applied; i.e., the video with proposed algorithm is more continuous, providing a predicted image similar to the original image to be updated.

However, the algorithm cannot deliver the change of environment instantly, as reflected in Figure 9c and the video with a moving subject. The motion of a moving subject is not reflected until the image is updated in (3). This is a fundamental limitation of the algorithm, because the algorithm utilizes delayed images for the prediction. Also, inappropriate margins negatively affect user immersion as we concerned in Section III, and as shown in the last part of the supplementary video and Figure 10, limiting the user-side FOV. Thus,
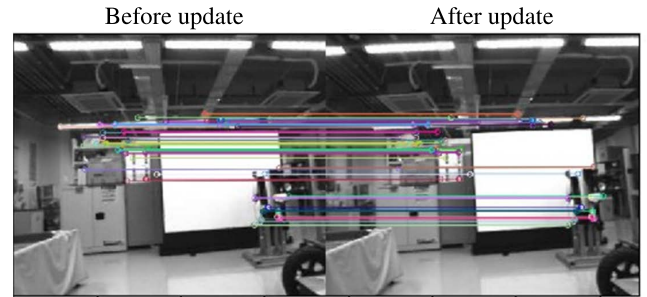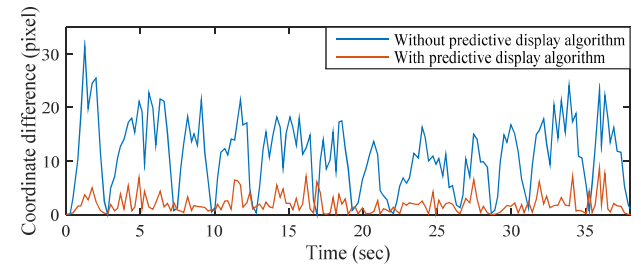
the margins must be selected to satisfy the criteria introduced in Section III.

V. CONCLUSION

In this research, a display system with a predictive display algorithm was developed. A robot neck-camera module was designed and manufactured based on a kinematic model of the human head-neck complex. The image delay during the tele-operation was analyzed; this may trigger simulator sickness. To deal with this issue, a predictive display algorithm was developed based on the camera geometry and human and robot kinematics. The image for the current user head orientation was predicted and provided to the user, exploiting the difference between the current user head orientation and the delayed camera orientation. Thus, the time-varying delays in communication and operation were compensated, and continuous images were thus provided to the user. The proposed algorithm is simple to implement in real-time tele-operation systems without expensive sensors or computational burden, yet powerful as demonstrated in the experiments. The feature analysis result shows that the proposed algorithm is effective in predicting images.

However, some challenges remain. The modeling uncertainties of camera geometry, such as tangent approximation and certain aspects of the pinhole camera model, should be reflected in the algorithm. The cropping margin decision strategy requires refinement for use under various telecommunication network conditions. If the user's motion is fast enough to escape the margins, the proposed algorithm cannot provide properly predicted image, as in [35]. Also,

the predictive display algorithm does not instantly reflect a change in the environment, because the algorithm used delayed images. Nevertheless, the proposed algorithm is a significant practical utility because it is simple and imposes a low computational burden. User evaluation in various situations is essential; we will quantitatively evaluate improvements in user comfort and employ these data to guide our future research.

REFERENCES

[1] Y. Jung, "A wearable control interface for Tele-operated robots." Doctoral thesis, Ulsan Nat. Inst. Sci. Technol., Ulsan, South Korea, Feb. 2019.

[2] L. Bevere, R. Sharan, S. Barrett, and C. Honegger, "Natural catastrophes and man-made disasters in 2016: A year of widespread damages," *Sigma*, vol. 2, pp. 1–43, Feb. 2017.

[3] D. Guha-Sapir, P. Hoyois, P. Wallemacq, and R. Below, "Annual disaster statistical review 2016: The numbers and trends," Centre Res. Epidemiol. Disasters, Brussels, Belgium, Oct. 2017, pp. 1–80.

[4] J. Lim *et al.*, "Robot system of DRC-HUBO+ and control strategy of team KAIST in DARPA robotics challenge finals," *J. Field Robot.*, vol. 34, no. 4, pp. 802–829, 2017.

[5] S.-J. Yi, S. G. McGill, L. Vadakedathu, Q. He, I. Ha, J. Han, H. Song, M. Rouleau, B.-T. Zhang, D. Hong, M. Yim, and D. D. Lee, "Team THOR's entry in the DARPA robotics challenge trials 2013," *J. Field Robot.*, vol. 32, no. 3, pp. 315–335, 2015.

[6] M. Johnson *et al.*, "Team IHMC's lessons learned from the DARPA robotics challenge: Finding data in the rubble," *J. Field Robot.*, vol. 34, no. 2, pp. 241–261, 2017.

[7] S. Gray, R. Chevalier, D. Kotfis, B. Caimano, K. Chaney, A. Rubin, and K. Fregene, "An architecture for human-guided autonomy: Team TROOPER at the DARPA robotics challenge finals," *J. Field Robot.*, vol. 34, no. 5, pp. 852–873, 2017.

[8] C. L. Fernando, M. Furukawa, T. Kurogi, S. Kamuro, K. Sato, K. Minamizawa, and S. Tachi, "Design of TELESAR V for transferring bodily consciousness in telexistence," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2012, pp. 5112–5118.

[9] M. Schwarz, T. Rodehutskors, D. Droeschel, M. Beul, M. Schreiber, N. Araslanov, I. Ivanov, C. Lenz, J. Razlaw, S. Schüller, D. Schwarz, A. Topalidou-Kyniazopoulou, and S. Behnke, "NimbRo rescue: Solving disaster-response tasks with the mobile manipulation robot momaro," *J. Field Robot.*, vol. 34, no. 2, pp. 400–425, 2017.

[10] S. Park, Y. Jung, and J. Bae, "An interactive and intuitive control interface for a tele-operated robot (AVATAR) system," *Mechatronics*, vol. 55, pp. 54–62, Nov. 2018.

[11] J. Golding and H. Markey, "Effect of frequency of horizontal linear oscillation on motion sickness and somatogravic illusion," *Aviation, Space, Environ. Med.*, vol. 67, no. 2, pp. 121–126, 1996.

[12] E. L. Groen and J. E. Bos, "Simulator sickness depends on frequency of the simulator motion mismatch: An observation," *Presence, Teleoperators Virtual Environ.*, vol. 17, no. 6, pp. 584–593, 2008.

[13] M. E. S. Pierre, S. Banerjee, A. W. Hoover, and E. R. Muth, "The effects of 0.2 Hz varying latency with 20–100 ms varying amplitude on simulator sickness in a helmet mounted display," *Displays*, vol. 36, pp. 1–8, Jan. 2015.

[14] G. Niemeyer and J.-J. E. Slotine, "Using wave variables for system analysis and robot control," in *Proc. IEEE Int. Conf. Robot. Autom.*, Apr. 1997, pp. 1619–1625.

[15] G. Niemeyer and J.-J. E. Slotine, "Stable adaptive teleoperation," *IEEE J. Ocean. Eng.*, vol. 16, no. 1, pp. 152–162, Jan. 1991.

[16] D. Lee and M. W. Spong, "Passive bilateral teleoperation with constant time delay," *IEEE Trans. Robot.*, vol. 22, no. 2, pp. 269–281, Apr. 2006.

[17] A. Aziminejad, M. Tavakoli, R. V. Patel, and M. Moallem, "Transparent time-delayed bilateral teleoperation using wave variables," *IEEE Trans. Control Syst. Technol.*, vol. 16, no. 3, pp. 548–555, May 2008.

[18] D. Sun, F. Naghdy, and H. Du, "Wave-variable-based passivity control of four-channel nonlinear bilateral teleoperation system under time delays," *IEEE/ASME Trans. Mechatronics*, vol. 21, no. 1, pp. 238–253, Feb. 2016.

[19] D. Sun, F. Naghdy, and H. Du, "Neural network-based passivity control of teleoperation system under time-varying delays," *IEEE Trans. Cybern.*, vol. 47, no. 7, pp. 1666–1680, Jul. 2017.

[20] C. Yang, X. Wang, Z. Li, Y. Li, and C.-Y. Su, "Teleoperation control based on combination of wave variable and neural networks," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 8, pp. 2125–2136, Aug. 2017.

[21] Z. Chen, Y.-J. Pan, and J. Gu, "Integrated adaptive robust control for multilateral teleoperation systems under arbitrary time delays," *Int. J. Robust Nonlinear Control*, vol. 26, no. 12, pp. 2708–2728, 2016.

[22] H. Wang, P. X. Liu, and S. Liu, "Adaptive neural synchronization control for bilateral teleoperation systems with time delay and backlash-like hysteresis," *IEEE Trans. Cybern.*, vol. 47, no. 10, pp. 3018–3026, Oct. 2017.

[23] P. M. Kebria, A. Khosravi, S. Nahavandi, P. Shi, and R. Alizadehsani, "Robust adaptive control scheme for teleoperation systems with delay and uncertainties," *IEEE Trans. Cybern.*, to be published.

[24] Z. Chen, F. Huang, W. Chen, J. Zhang, W. Sun, J. Chen, J. Gu, and S. Zhu, "RBFNN-based adaptive sliding mode control design for delayed nonlinear multilateral tele-robotic system with cooperative manipulation," *IEEE Trans. Ind. Informat.*, to be published.

[25] K. Theofilis, J. Orlosky, Y. Nagai, and K. Kiyokawa, "Panoramic view reconstruction for stereoscopic teleoperation of a humanoid robot," in *Proc. IEEE-RAS 16th Int. Conf. Hum. Robots*, Nov. 2016, pp. 242–248.

[26] M. Jagersand, A. Rachmielowski, D. Lovi, N. Birkbeck, A. Hernandez-Herdocia, A. Shademan, D. Cobzas, and K. Yerex, "Predictive display from computer vision models," in *Proc. Int. Symp. Artif. Intell., Robot. Automat. Space*, 2010, pp. 673–680.

[27] R. Azuma and G. Biship, "A frequency-domain analysis of head-motion prediction," in *Proc. SIGGRAPH*, vol. 95, 1995, pp. 401–408.

[28] J. R. Wu and M. Ouhyoung, "On latency compensation and its effects on head-motion trajectories in virtual environments," *Vis. Comput.*, vol. 16, no. 2, pp. 79–90, 2000.

[29] W. S. Kim and A. K. Bejczy, "Graphics displays for operator aid in telemanipulation," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 1991, pp. 1059–1067.

[30] T. Yuichi and M. Uchiyama, "Predictive display of virtual beam for space teleoperation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 1996, pp. 1544–1549.

[31] F. Richter, Y. Zhang, Y. Zhi, R. K. Orosco, and M. C. Yip, "Augmented reality predictive displays to help mitigate the effects of delayed telesurgery," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 444–450.

[32] M. J. Brudnak, "Predictive displays for high latency teleoperation," in *Proc. NDIA Ground Vehicle Syst. Eng. Technol. Symp.*, 2016, pp. 1–16.

[33] Oculus VR. (2018). *Oculus Rift*. [Online]. Available: https://www.oculus.com/

[34] FLIR. (2018) *Flea3*. [Online]. Available: https://www.flir.com/

[35] E. C. Edwards, "Method and system for time/motion compensation for head mounted displays," U.S. Patent 7 312 766 B1, Dec. 25, 2007.

YEONGTAE JUNG received the B.S. degree in mechanical and advanced materials engineering and the Ph.D. degree in mechanical engineering from the Ulsan National Institute of Science and Technology (UNIST), Ulsan, South Korea, in 2013 and 2019, respectively. He is currently a Senior Researcher with Korea Shipbuilding and Offshore Engineering, Ulsan. His current research interests include modeling, design and control of human–robot interaction systems, teleoperation systems, and actuation mechanisms. He received the Korean Government Minister Awards from the Ministry of Science, ICT and Future Planning, in 2017. He was a recipient of the Global Ph.D. Fellowship supported by the National Research Foundation of Korea, from 2013 to 2018.

KYUTAEK HAN received the B.S. degree in mechanical system design engineering from the Seoul National University of Science and Technology, Seoul, South Korea, in 2016, and the M.S. degree in mechanical engineering from the Ulsan National Institute of Science and Technology (UNIST), Ulsan, South Korea, in 2019. He is currently a Researcher with the Korea Culture Technology Institute. His current research interest includes imaging technology and platform for immersive experience.

JOONBUM BAE (S'07–M'12) received the B.S. degree *(summa cum laude)* in mechanical and aerospace engineering from Seoul National University, Seoul, South Korea, in 2006, and the M.S. degree in mechanical engineering, the M.A. degree in statistics, and the Ph.D. degree in mechanical engineering from the University of California at Berkeley, Berkeley, CA, USA, in 2008, 2010, and 2011, respectively.

In 2012, he joined the Department of Mechanical Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan, South Korea, where he is currently the Director of the Bio-Robotics and Control (BiRC) Laboratory. His current research interests include modeling, design, and control of human-robot interaction systems, soft robotics, and biologically inspired robot systems.

Dr. Bae received Korean Government Minister Awards from the Ministry of Public Safety and Security and the Ministry of Science, ICT and Future Planning, in 2016 and 2017, respectively. He also received the Young Researcher Award from the Korea Robotics Society, in 2015. He was a finalist of the semi-plenary Paper Award at the American Society of Mechanical Engineers Dynamic Systems and Control Conference, in 2012, and a finalist of the Best Poster Paper Award at the International Federation of Automatic Control World Congress, in 2008. He was a recipient of a Samsung Scholarship during his Ph.D. studies.

• • •