1. Introduction

Nowadays, remote controlled robots play a decisive role in situations and environments that are dangerous, unstructured, and under-recognized, such as bomb disposal, rescue, or space exploration as shown in Fig. 1, (Yamauchi, 2004; Keiji et al., 2011; Diftler et al., 2011). These roles are accomplished due to the robotic remote control technique that combines the intelligence of humans with the abilities of robots (e.g., precision and mechanical strength). However, human operators still frequently deal with a lack of intuitive control interfaces. The absence of intuitive human-robot interfaces is caused by a lack of proper feedback and a complex motion command input method, which causes decreased task efficiency and may lead to failure. To overcome this issue, various studies have been completed in recent decades to improve the intuition of the human-robot remote interface.

Meanwhile, the immersive 3D virtual collaboration system has been studied thoroughly as a means of communication, and much progress has been made in recent years. Some works have attempted to provide a virtual experience in a complex remote world for high-level communication. Although it is possible to make people feel like they are in a remote place with realistic feedback devices, there is still a limitation to what people can do in the absence of physical connection. If we can make the physical effects of the remote object real through this virtual collaboration over the internet, the distance restraint will no longer be a problem. Fortunately, the robotic remote control system is capable of satisfying this desired task. The virtual collaboration system can make this robotic remote control system more realistic so that human operators can perform effectively with intuitive robot control (e.g., Monferrer and Bonyuet, 2002; Yamada et al., 2008; Zang et al., 2016).
Fig. 2 shows the basic structure of the robotic remote control methods that apply the virtual collaboration system. The virtual environment is connected to each remote real site (an operator side and a robot side) and the remote sites give and receive environmental information, robot position command, etc. Therefore, the virtual environment acts as a medium that connects the remote sites. There are two main advantages to linking the remote sites through the virtual environment. First, we can substantially reduce the amount of data translation because it is possible to obtain visual feedback and other various multimodal feedback data from a local site based on the virtual environment. Consequently, the real-time performance can be significantly improved. Second, it is controllable from the third person viewpoint, and many operators can participate simultaneously from a multilateral viewpoint in virtual space, which leads to efficient and safe operation.

Thus, the robotic remote control using the virtual collaboration scheme is able to contribute where transmission data loss or delay often occurs, such as remote control over a long distance (e.g., between continents or earth to space). Also, it can greatly increase the efficiency of operation performance in areas where it is difficult to obtain clear visibility, such as a remotely operated underwater vehicle (ROV). For this robotic remote control via virtual collaboration, the following important element technologies are necessary.

- Remote control based on human motion
- Multimodal feedback scheme
- Virtual collaboration for remote control

![Diagram](image)

Fig. 2 Structure of robotic remote control based on human motion via virtual collaboration
In the case of motion capture technology with remote control based on human motion, it is necessary to guarantee real-time performance, since the motion command trajectories of each robot joint are created based on the human motion. Therefore, frame rate and accuracy are the main issues, in addition to the motion capture device not disturbing the natural movement of the human operator and stable acquisition of the motion. In the case of remote control, maintaining high stability in a situation with a time delay is a major problem. When passing human movements to the remotely operated robot, it is necessary to consider how accurately the robot moves with the human motion and the joint coordinate information transfer to the robot, which has a different joint structure than the human.

For multimodal feedback, how well the system can provide a realistic immersive feeling is important, since it is a real interaction with the remote environment. Therefore, techniques for enhancing perception are required. Mobility of operating devices has been recently emphasized, and portable feedback equipment have been tested for giving a satisfactory immersive feeling. How to reconstitute the virtual environments of the remote sites for the virtual collaboration area is an important part and various methods have also been tried to provide the virtual space for the human operator (e.g., 3D screen, cave, virtual reality (VR) or augmented reality (AR) with head mount display (HMD)).

This paper focuses on the implementation parts of previous research and the advantages when they are applied to robotic remote control. The arrangement of this paper is as follows. Previous research on robotic remote control based on motion capture is discussed in Section 2. Previous research on the multimodal feedback scheme is discussed in Section 3. In Section 4, previous research on virtual collaboration with remote control is detailed. Finally, Section 5 concludes the paper by summarizing and discussing the potential of robotic remote control with virtual collaboration system.

<table>
<thead>
<tr>
<th>Method</th>
<th>Advantage</th>
<th>Limitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exoskeleton mechanical devices</td>
<td>High accuracy</td>
<td>Too sophisticated</td>
</tr>
<tr>
<td></td>
<td>Low cost sensors</td>
<td>Disturb natural movements</td>
</tr>
<tr>
<td></td>
<td>Provide force feedback</td>
<td>No global position</td>
</tr>
<tr>
<td>Fiber optics sensors</td>
<td>Robust from noise</td>
<td>Less accuracy</td>
</tr>
<tr>
<td></td>
<td>Finger motion detection</td>
<td>Low durability</td>
</tr>
<tr>
<td></td>
<td>Daily clothes</td>
<td>High-maintenance</td>
</tr>
<tr>
<td>Electromagnetic</td>
<td>High accuracy 6 DOF</td>
<td>Interference from metal objects</td>
</tr>
<tr>
<td></td>
<td>Less occlusion</td>
<td>Limited range</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High cost</td>
</tr>
<tr>
<td>Inertial measurement unit (IMU)</td>
<td>Portable</td>
<td>Cumulative error</td>
</tr>
<tr>
<td></td>
<td>Natural movement</td>
<td>Drift error</td>
</tr>
<tr>
<td></td>
<td>No occlusion</td>
<td>Difficult to detect Finger motion</td>
</tr>
<tr>
<td>Marker-based optical system</td>
<td>High accuracy</td>
<td>Fixed and limited area</td>
</tr>
<tr>
<td></td>
<td>Flexibility</td>
<td>Controlled lighting</td>
</tr>
<tr>
<td></td>
<td>Natural movement</td>
<td>Occlusion</td>
</tr>
<tr>
<td>Markerless optical system</td>
<td>Natural movement</td>
<td>Less accurate</td>
</tr>
<tr>
<td></td>
<td>Low cost</td>
<td>Controlled lighting</td>
</tr>
<tr>
<td></td>
<td>Portable</td>
<td>Occlusion</td>
</tr>
<tr>
<td></td>
<td>Flexibility</td>
<td>Software complexity</td>
</tr>
</tbody>
</table>
2. Motion-based Remote Control

Because robots can replace humans in dangerous tasks, and can also offer accurate, robust, and fast performance for precise operation and repetitive tasks, remote control schemes have been studied to achieve a more natural and intuitive remote control environment. As a result, human motion is being used to manipulate remote controlled robots, instead of conventional devices, such as the control pad or joystick. Advances in remote control schemes have facilitated the development of motion capture methods. In the case of remote control that directly generates trajectories of robot joints based on human motion capture data, sufficient frame rate of motion capture is required for stable robot control. DeGoede et al. (2001) measured the movement speed of the human hand, and in the case of healthy adults, it was up to approximately 3.0 m/s. Therefore, assuming that the hand moves at a maximum of 3 m/s, the frame rate must be at least 150 Hz to measure the distance the hand moved at a resolution of 20 mm or less for each frame. Of course, this may differ depending on the motion capture methods. Depending on the system, it is often not necessary to move as fast as 3.0 m/s, so the required frame rate fluctuates according to the purpose of the remote robotic system. However, it is highly recommended to provide a high enough frame rate to precisely measure the movement of a person who moves swiftly, so the motion capture system must be designed and selected accordingly.

2.1 Motion capture technologies

A comparison of recent motion capture methods is presented in Table 1. Commonly used human motion capture interfaces can be divided into wearable types and non-wearable types. Examples of wearable motion capture methods are exoskeletal mechanical devices (e.g., Zhou and Ben-Tzvi, 2015; Mallwitz et al., 2015; Toyota, 2017), fiber optic sensors (e.g., Stupar et al., 2012; Koyama et al., 2018), inertial measurement units (IMU) (e.g., Roetenberg et al., 2009; Brigante et al., 2011; Seel et al., 2014), electromagnetic trackers (e.g., O’Brien et al., 1999), and electromyographic muscular activity sensors (Fukuda et al., 2003).

In the case of those wearable motion capture devices, the attached cables, sensors, and other devices can hinder natural and dexterous motion of humans and increase fatigue level. Therefore, researchers (e.g., Miura et al., 2005; Liu et al., 2011) have tried to reduce the number of wearable sensors. The non-wearable type is mostly implemented using the optical motion capture scheme, which can be divided into marker-based and markerless motion capture methods. The marker-based optical gesture capture systems compute the 3D position of a subject from captured images using single or multiple cameras. Data acquisition is traditionally implemented using special markers, such as reflective markers, attached to a human operator (e.g., Guerra-Filho et al., 2005; Kirk et al., 2005; Canton-Ferrer et al., 2010; Merriaux et al., 2017). These marker-based methods accurately detect the human operator motions, but sometimes the attached marker can interrupt natural movements. The markerless method uses the multi-view image and 3D skeletal model to capture the human motion (e.g., Menier et al., 2006; De Aguiar et al., 2007; Gall et al., 2009; Liu et al., 2011; Hasler et al., 2009). Ye et al. (2011) applied a depth camera to capture motion from the 3D point cloud. Some works (e.g., Bachmann et al., 2001; Corrales et al., 2008; Zihajehzadeh et al., 2015) merged various types of motion capture sensors, in order to leverage the advantages and disadvantages of each motion capture method.

![Motion capture methods](image_url)

Fig. 3 Human motion capture methods (a) exoskeletal mechanical device – Toyota; (b) marker-based optical motion capture – Canton-Ferrer et al.; (c) markerless optical motion capture – Gall et al.
2.2 Difficulties in kinematic discrepancy

The key difficulty of the motion-based remote control is kinematic discrepancy between the human and the robot. Since the kinematics of the human and the robot are different even in the case of humanoid robots, the captured motion should be processed before it can be applied to the robot. There are several methods for mapping the captured motion into the trajectory of the robots. In the early stage, the kinematic discrepancy was solved by constraining the captured motion data to match the robot kinematics (e.g., N. S. Pollard et al., 2002; A. Dasgupata et al., 1999). In this method, first, the captured motion data is mapped to the intermediate skeletal model whose kinematic structure is identical with the robot kinematics, and then joint angle and velocity limits are applied. Another approach for solving the kinematic discrepancy is to transform raw motion data to the motion primitives (e.g., M. Field et al., 2009). A range of stochastic models and sequencing algorithms such as Hidden Markov Model (HMM) (e.g., T. Inamura et al. 2004) and Factorial HMM (FHMM) (e.g., D. Kulic et al. 2008) have been applied to solve the problem. Another method is to transform high dimensional motion data to a low dimensional manifold using analysis such as Principal Component Analysis (PCA) or non-linear methods such as Isomap and Gaussian Process Latent Variable Model (GPLVM) (e.g., J. M. Wang et al., 2008; A. P. Shon et al., 2005).

There have been various practical studies to solve the kinematic discrepancy in motion-based remote control. From the early stage of motion-based robotic remote control, one or two parts of the human body, such as the hands or the head, was used to control end effectors or other parts of teleoperated robots. Using a part of the human body position data, which is obtained from a certain motion capture device, a part of a robot part is controlled after interpretation of the kinematics. Kofman et al. (2005) implemented an end-effector of a robot manipulator teleoperation control scheme based on the marker-based optical gesture capture method, which is shown schematically in Fig. 4(a).

In the case of the humanoid robots, it is comparably easy to map the joint position from the human skeleton information. However, wheel or crawler type mobile robots and complicated flexible or redundant robot manipulators are not able to map the joint position since the joint structure is different with human. Many works have tried to apply various methods to solve this issue. In fact, the robotic remote control has been applied to wheel or caterpillar track type mobile robots almost from the start. Most of these robots are still operated by steering wheels or joysticks, not by human motions to this day (e.g., Kelly et al., 2011; Sulaiman et al., 2015; Schwarz et al., 2017). But for the obstacle perception, more and more systems have applied the bilateral system which equipped parallel or serial manipulators for the motion capture. The human hand position is captured by the position of the end effector of those manipulators and it can be scaled and direct mapped to the center of the mobile robot body (e.g., Li et al., 2017). In the case of the crawler type mobile robot, the center of body is able to move freely in the 3D space. Thus, You et al. (2017) applies 3 degree of freedom (DOF) parallel manipulator type bilateral master controller.

Regarding the joint mapping for the complicated redundant manipulator, three different approaches have been studied. First one is the end effector position computation by the inverse kinematics based on the human hand motion detection (Gong et al., 2018; Xu et al., 2018). Second one is the mode change method. Schwarz et al. (2017) and Kikuchi et al. (2017) divide the tasks, such as lifting, handling and walking then use different types of motion capture devices for each task by fixing others. Last one is multi master method. The parts of the redundant manipulators are separately controlled by multiple human operators control for reducing the task complexity (e.g., Malysz et al., 2011). Especially it is difficult to map the human motion to the soft manipulator. Still, most works using end effector control by human hand motion detection. Li et al. (2015) controlled the end effector based on human hand motion by controlling stiffness of the part of the manipulator for the more dexterous manipulation. In other to have the merits of the soft manipulator, the intuitive automatic joint mapping method is essential. Sanna et al. (2015) proposed automatic mapping method by exploiting graph similarity criteria, user preferences, motion constraints and information about the armature topology.

The kinematic discrepancy can lead singular configuration to the remotely controlled robot. Rakita et al. (2017) proposed an algorithm that can avoid the singular point and perform motion smoothing in real time to overcome the limitations when mapping the position command of the robot end-effector directly from the human hand motion. In order to deal with more complicated and diverse objects, there are many successful works presented to control the dual-armed robots remotely. As shown in Fig. 4(b), Kruse et al. (2015) controlled the dual-arm robot based on the human arm motion captured by the depth camera. Nicolis et al. (2018) proposed the occlusion-free scheme using visual servoing with the camera, which was placed on the end effector to control both arms to reduce operator fatigue. These dual arm remote control schemes have been actively applied to surgical robots in the last decade (e.g., Ballantyne, 2002; Ferraguti et al., 2002).
Unlike the above studies that control the robot gripper remotely with human motion to perform physical work, to provide more natural and intuitive visual and other feedbacks, attempts to control the attitude of the camera or other sensors and the overall links of the robot based on human head or torso motion are more prevalent. Watanabe et al. (2008) introduced the TORSO-HMD system which can express the upper body motions of humans to acquire natural and comfortable visual information. Recently, researchers applied the whole-body gestures to humanoid-type robots in order to handle daily objects and environments that are highly adapted to humans. Although it is not a real-time teleoperation, Shiratori et al. (2007) developed a humanoid robot that can perform very complex movements, such as a dance performance, using motion capture. Kim et al. (2009) proposed a framework of walking imitation between a human and a humanoid robot. Using IMU, walking motions of humans are recognized and used as humanoid robot control inputs for on-line walking imitation. Lee et al. (2010) developed an upper body teleoperation using marker-based optical gesture capture system. This work segments a sequence of the human operator motion and this can be useful for a system to autonomously learn and develop through human movement imitation (see Fig. 5).

It is a very challenging task to achieve the stable position control of the whole robot body including both the upper body and the lower body, taking the same posture as the operator. The robust posture stability of the robot must be implemented to simultaneously achieve the same posture as the operator. In order to solve this problem, Koenemann et al. (2014) proposed a method of ensuring stability with a system where a remotely controlled robot finds a stable center of gravity of all attitudes locally and retargets the attitude in real time (see Fig. 6(a)). As shown in Fig. 6 (b), Otani and Bouyarmane (2017) showed how to control dynamic partitions by distinguishing the fixed support contact, the manipulation contract, and the contact-free tracking links from each link of the robot in order to ensure stable operation of the remotely placed robot with the physical manipulation in the remote environment.

Fig. 4 Structure of the motion based remote control (a) end-effector control using marker-based motion capture – Kofman et al.; (b) dual arm control using markerless motion capture – Kuruse et al.

Fig. 5 Humanoid robot remote control based on human motion. (a) complex movement control by prerecorded human motion – Shirator et al.; (b) real time control using the human motion captured by the IMU (Kim et al.(b)) and the optical sensor (Lee et al.(c))
3. Multimodal feedback Scheme

Teleoperation systems started with the use of unilateral control, or one-way information communication from a human operator to a robot. Later, the bilateral method was introduced, which can offer feedback, such as force feedback, to the human operator from the robot side (Minsky, 1980). Recently, to realize a more realistic human-robot interface, the multimodal feedback method was proposed, and this can be achieved using both force feedback and variable feedback, such as force, tactile, visual, and sound. Tachi et al. (2011) presented the multi-modal multi-user telepresence and teleaction system. This study offered visual, force, and 3D sound to a human operator. Hand motion is captured by the stationary human-system interface or mobile human-system interface, which is also used for force feedback. Some researchers have applied a novel force feedback interface to achieve a far better sense of reality. Arias and Hanebeck (2009) used a novel force feedback interface to implement extended workspaces (see Fig. 8(a)); however, for precise force feedback rendering, a large portal carrier system has to be installed with ample space. Fig. 8(b) shows the FlexTorque, which is similar to the human muscle structure, and can offer force feedback using simple devices (Tsetserukou et al., 2010). As shown in Fig. 7(c), one haptic interaction scheme using a string-based interface called the Space Interface Device for Artificial Reality (SPIDAR) applies multiple strings for force feedback (Sato, 2002). In addition, Yin et al. (2016) designed a method to transmit the force feedback, which is received by the robot from a remote place to the operator by using magnetorheological (MR) fluids to convert the viscosity of the fluid into an electric signal.

Force feedback schemes, as described above, are intuitive and realistic because the operator can feel and interact with the force directly from the robot site, so that precise work is possible and even beginner operators can easily adapt. However, the force feedback system often has unrealistic cases. Force feedback, therefore, would require a significant amount of force to be applied to the operator in certain situations, such as when the operator is pushing hard against a wall. This requires expensive and massive equipment, and there are also significant safety considerations. As an alternative, researchers are focusing on vibratory feedback (e.g., Bloomfield and Badler, 2008). Using small vibratory motors along the arm and hand, tactile feedback can be presented to the operator without the disadvantages of full force feedback. As shown in Fig. 8(a) Pacchierotti et al. (2016) proposed a tactile feedback method by applying two motors, DC pressure and AC pressure (vibrations), to the fingers of the operator, so that the tactile information on the tool tip of the surgical robot can be more realistically fed back. In these kinds of methods using vibration motors, an actuator that can be driven with a wide range of frequencies is necessary to provide a much better subdivided sensation to an operator. Yang et al. (2016) implemented a small actuator that can oscillate a wide frequency band using unstable mass. Adel et al. (2017) implemented force feedback by controlling the magnetic fields and mapping the magnetic force between the magnet placed on fingers and the electromagnetic plate. Although there are drawbacks that it can be used only in a limited area, the range of force that can be passed is wide, thus it is possible to implement more precise resolution in the future, and by wearing relatively light equipment, operators can feel more natural and accurate environmental information from the robot site.
applied in the future, is the method using a soft actuator. Kato et al. (2007) made a thin flexible tactile actuator array with a thickness of 1 mm. When power is supplied, each polymeric actuator (diameter approximately 2 mm) is able to operate to a displacement up to 0.4 mm with a force of 1.5 gf on a rate of 2 Hz. Ishizuka et al. (2014) proposed a scheme that can change the stiffness of a diameter smaller than 5 mm by deforming the pressure in the range of 200 to 600 kPa as a method of forming a magnetic field at a target site using the MR fluid and increasing the density of the iron particle of the MR fluid. Phung et al. (2017) presents a soft actuator with a diameter of 130 mm that can drive a wide range of 0 to 150 Hz rate with a force up to 50 mN using a piezo element. The above research can enhance the feedback reality further by applying wearable gloves and the other devices. This leads to the improvement of the remote operation task performance. Also, Kimura et al. (2014) proposed a softness feedback method of an object by using a linear voice coil motor.

Researches on providing force or tactile feedback without physical contact are also ongoing. Carter et al. (2013) and Long et al. (2014) implemented the system to deliver haptic feedback through the air in real time using a focused ultrasound made with ultrasound transducer 2D arrays (see Fig. 8(b)). Gupta et al. (2013) proposed a system that transmits the air vortex ring of different sizes and strengths up to several meters by using the linear motion characteristics of the air vortex ring (see Fig. 8(c)). Hashizume et al. (2017) implemented a device that can combine focused ultrasound and air vortex to enable tactile feedback of a finer resolution over a wider range. Since there is no physical contact, the amount of the force feedback transfer is limited, and there are many differences in precision depending on the distance, but it can be used immediately without wearing different equipment, it can be applied to remote control with various forms.

Also, over the past few years, the visual feedback system has evolved with the growth of quantity and quality 3D contents. At present, a considerable number of devices, such as HMD and 3D television, has also spread to ordinary households, and is no longer limited to industrial or professional fields. In fact, these visual feedback devices (e.g., 3D screen, cave, HMD, transparent HMD) are closely related to the next section, virtual collaboration system, so they are detailed in the next section.

In addition, it is also conducted to feedback the temperature detection of the remote area to the operator and to transfer the sense of the remote place more realistically. Peiris et al. (2017) applied five thermal modules to the HMD. Each thermal module is controlled with a ±3 °C/s rate. This work shows how people can efficiently find a heat source and feel the weather with this presented HMD. Mehra et al. (2015) introduced an interactive sound propagation scheme based on the wave equation which can accurately generate the sound for dynamic and directional sources. Manghisi et al. (2017) presented a 3D audio feedback system using the speaker arrayed room, called the MATE. This room can also change temperature and humidity control for better realistic remote environment feedback.
Researchers for many years have been studying virtual face-to-face schemes for an effective and close collaboration. Such efforts have resulted in the conference call or the video conferencing systems, which are already commercially available, and the most successful examples of the virtual collaboration system. Recently, in order to offer physical interactions between remote places via virtual collaboration, these types of virtual collaboration systems are currently applying teleoperation and have developed the telepresence concept (Minsky, 1980). In the case of the medical field, Ellison et al. (2004) showed how the virtual collaboration system is important and useful to interact between medical doctors and patients using an LCD screen mounted on a mobile robot. Neustaedter et al. (2016) showed how efficiently collaboration can be between remote people using this kind of mobile robot based on the virtual collaboration system shown in Fig. 9(a). However, such a system lacks DOF and actuators to communicate physically. Fernando et al. (2012) introduced an advanced telepresence system called TELESAR V (see Fig. 9(b)). This system uses the 52 DOF humanoid robot to transfer the accurate movements of the remote operator, offers immersive stereo vision on the side of the operator, and has a human-like arm actuator to perform realistic interaction. But these kinds of telepresence systems need a stable and fast network as an essential condition because those systems transfer a large amount of data in real time (e.g., high-definition stereo video stream, force or tactile information, position control data, etc.). Therefore, applying this kind of system is not suitable where the network condition is unstable (e.g., time delay, data loss, or limited amount of data transfer), such as intercontinental or interspace long distance remote control.

4.1 Virtual collaboration for communication time delay

In the remote control, the time delay and data missing problem are inevitably generated since the motion or position command data and the feedback data between the remote places are transferred via the network communication. The problem of time delay is that on the slave side, continuous position input is not conveyed and can cause problems of stability during a certain task execution. On the master side, the feedback signal may be unnatural for the operator, so the remote operation task efficiency decreases and can even make a false determination. Islam et al. (2015) showed stability and tracking control problems of remote control under the conditions with time-varying delay in various scenarios. Shokri-Ghaleh and Alfì (2014) provided the results of comparing the performance of various remote control optimization algorithms in environments with time delay and model uncertainty. Hua et al. (2015) estimated the lost velocity command of robot links using a sliding-mode velocity observer, and designed the saturated proportion and saturated damping controller to compensate for the uncertainty of the model based on the estimated velocity input information. Rank et al. (2016) proposed a way to regenerate position input data by using the probability prediction model data when data is lost or time delay occurs. In spite of the efforts stated above, in systems where the distance is very far or it is hard to ensure a stable network condition (e.g., where missing data can occur frequently), we need to figure out and apply a method to reduce the amount of data for the stable and sufficient network traffic condition.
In order to deal with the time delay, there are two ways. One is the prevention method and another one is the interpolation method when the time delay occurs. Firstly, to prevent the time delay, the network access protocol optimization algorithm is proposed by Condoluci et al. (2017). However, to prevent the time delay fundamentally, transmission data reduction is basically necessary. In fact, the visual feedback data occupies the majority of the data, thus, visual feedback substitution schemes by using the virtual reality based are emerging which are detailed in Section 4.

Secondly, if the time delay occurs, the remote control systems are unstable both the master and slave side. Hua et al. (2015) estimated the lost velocity command of robot links using a sliding-mode velocity observer, and designed the saturated proportion and saturated damping controller to compensate for the uncertainty of the model based on the estimated velocity input information. Rank et al. (2016) proposed a way to regenerate position input data by using the probability prediction model data when data is lost or time delay occurs. To secure the stability, the Lyapunov-Krasovskii functional method is applied for the filtering of the error (e.g., Zhai et al., 2017; Hua et al., 2017). And recently, thanks to the advancement of computing power, there are many works to interpolating the control input and trajectory by using neural networks (e.g., Li et al., 2016; Wang et al., 2017; Sun et al., 2017). In spite of the efforts stated above, in systems where the distance is very far or it is hard to ensure a stable network condition (e.g., where missing data can occur frequently), we need to figure out and apply a method to reduce the amount of data for the stable and sufficient network traffic condition.

Therefore, researchers have been striving for stable remote control in situations where these time delays are present. Tidoni et al. (2017) proposed an algorithm to improve remote operation task efficiency by synchronizing sound feedback (from footstep of the robot) with visual feedback, in order to overcome the time delay of data occurring while performing remote control of the humanoid robot between Japan and Italy. When a time delay occurs, a video signal (about 800 ms delay occurred in this work) and sound signal do not normally occur with the same delay, so that the transferred video and sound signal are not matched. In this work, therefore, it was proven through experiments that the operator experienced confusion from time to time to perform the operation, the work efficiency dropped, and precise work performance became difficult. If the synchronized sound and video feedback is provided to the operator, even though there is a delay in the feedback, it can be learned so that more efficient work can be performed under the asynchronous feedback condition. Fig. 9(c) shows the actual remote control task using this system. However, in the case of this study, we could not identify and resolve the fundamental cause of the time delay. In fact, the time delay is closely related to the capacity of the data to be transmitted. Video signals for visual feedback especially occupy a significant portion of the transmitted data.

Researchers have introduced the method to apply VR to remote control due to the optimized video signal, which has been used in visual feedback over the past decade. Guo et al. (2016) reduced the amount of data transfer by providing the VR-based visual feedback to operators instead of using the real time streamed camera images on the long-range surgical robot system. As a result, the time delay of remote control between Japan and China was dramatically reduced.
to about 20 s. Xu et al. (2016) also proposed a way to overcome time delay by using visual feedback based on VR instead of video signals to reduce the amount of data. In this research, the applied VR is automatically generated by real-time 3D mapping using the depth camera placed on the robot side, instead of using a manually reconstructed 3D model. The generated VR 3D map is updated to the remote side operator only when new information is generated, so it is possible to construct a more realistic and practical optimized VR environment. Also, Huijun et al. (2007) proposed the force feedback rendering method on the operator side using the constructed 3D map. It is able to reduce the amount of force feedback data transferred to the operator over the long distance. By using VR in this manner, it is possible to reduce the amount of data by providing appropriate feedback based on the virtual environment local to the operator. Thus, if there is a time delay, it can prevent the phenomenon of decreasing reality, so remote operation efficiency can be improved.

Furthermore, with the VR integrated remote control method, there are also successful research cases using the advantage that a third party can collaborate more efficiently in the remote control operation by having wide viewpoints. The Virtual Collaboration Arena (VirCA) is a good example, which is presented by Galambos et al. (2014). This work consisted of some components that can be connected to each other according to the needs of users. The main component is the 3D visualization, which can be the bridge between the user and the virtual reality, and it allows the operators to interact with the virtual reality and its participants (see Fig. 10(a)). In this kind of system, the device and human operators connect the real world to the virtual one, so they can have an effect on each other. Baker et al. (2017) introduced Robonaut 2, which performs various tasks on behalf of people in outer space. Since it is necessary to perform dangerous work, in order to make a rational judgment, the VR environment provides many operators the ability to confirm the progress of work from various viewpoints (see Fig. 10(b)). In addition, by conducting a preliminary simulation verification, safer work is possible. As shown in Fig. 10(c), the Augmented Reality Interface for Teleoperation (ARITI) project applies augmented reality to help human operators to achieve complex tasks, such as cooperative teleassistance (Domingues et al., 2009). Multiple operators control the virtual phantom robot to take virtual objects as a simulation. After the validation, the distant real robot will follow the phantom and take the real objects. Osaki et al. (2009) conducted a study to share real-time 3D sketches using augmented reality. The concept of this research is easily applied to remote control and can be used like exchanging opinions when executing simulation before actual remote operation by multiple operators. Hou and Watanuki (2012) proposed a method to evaluate and verify the similarity of the work performance in the VR environment and in the actual working environment by comparing the brain activities during lathe work. By applying the virtual collaboration system to the remote control, as described above, it is possible to see the progress of the remotely operated task situation in various fields of vision more freely and it is possible to reasonably make a decision. In addition, verification through pre-simulation can be easily implemented. Therefore, it tends to be applied to various complicated and dangerous work, including the fields of medicine and space.

![Fig. 10](https://example.com/fig10.jpg)

**Fig. 10**  Robotic remote control with virtual collaboration systems. ((a) Galambos et al.; (b) Baker et al.; (c) Domingues et al.,)

### 4.2 Challenges in virtual collaboration

There are several challenges to implement high fidelity virtual collaboration system (e.g., C. P. Kuan et al., 2003). First, implementation of the telepresence of remote environments including visual, auditory, and haptic feedback is major
difficulties of the virtual collaboration. The method for reconstructing geometric information of remote environment, real-time haptic rendering algorithms, and real-time deformable object simulations have proposed to construct virtual remote environment as real as possible. Second, the I/O devices to deliver sensation of virtual environment to the user, and this issue were reviewed in Section 3. Third, inconsistency between virtual environment and remote environment. In the virtual collaboration system, the information of remote environment such as position of the robot, reaction force, geometry of the environment does not be given in real-time because of the communication delays. Thus, it is difficult to maintain consistency between virtual environment and the remote environment. Most of previous studies solve the inconsistency problem using physical model of the remote environment. If the physical model can simulate the remote environment perfectly, then it can guarantee stable and transparent remote control of the robot. However, it is still open problem to implement precise physical model of the remote environment. There are several difficulties to implement accurate model of the remote environments. First, the interaction data gathered in the remote environment possesses a lot of noises. Second, it is difficult to model interaction occurred in the remote environment because it involves quiet complex physical process. Third, it is necessary to update the model in real time once it recognizes the errors in the model. In previous, many types of real-time update algorithm such as sliding-average least-square algorithm (e.g., H. Li et al. 2007), dynamic identification of remote environment (e.g., A. Haddadi et al. 2008), parameter identification methods (e.g., T. Yamamoto et al., 2008), and modeling based on Kelvin-Voigt model (Z. Jiang et al. 2009) have been proposed. The other researches related this issue can be found in the literature (e.g., A. Achhammer et al., 2010; R. Schindeler et al., 2016). Most of the previous methods focused on the interaction modeling between robot and static environment, but there were little studies aiming practical interactions involving interaction with moving objects.

5. Conclusion

Progress in artificial intelligence (AI) increases the likelihood that fully autonomous robots will replace repetitive, time-intensive tasks for a human being. Robots are increasingly completing integral and collaborative tasks in a variety of fields. However, in highly unstructured dynamic environments where objects are unfamiliar, changing shape, or moving with unknown motion, human intelligence is needed to make decisions and control robots, especially in delicate and dangerous environments with complex operations. This demand has brought many advances in the robotic remote control area. There are many successful works, including some that are commercially available.

Recently, researchers have tried various schemes in order to offer a far better sense of reality and robustness to support intuitive operation. For the intuitive remote operation, many successful efforts apply part or all of the human body motion. As verified through several studies, if intuitiveness is high for remote control, it is possible for the operator to respond quickly in a dynamic and unfamiliar environment. The degree of fatigue during the remote operation task is also small, and it leads to the stability of task performance with precise work, and relatively high concentration is maintained, even after working for a long time. In addition, it is possible to reduce the training time for the expert robotic remote control operator, and there is merit to such a cost reduction.

Motion-based remote control is a key technology for the intuitive remote interaction with remotely placed robots. For that, we need to first apply the motion capture method to translating position command information on the robot. Because there are advantages and disadvantages for each motion capture method, the optimally suitable method for that system should be applied. Also, in the case of remote control using human motion, information on the joint position is used for robot control in real time, so researches are actively being carried out to maintain high stability in a situation with time delay issues.

For improved sensing of reality in the remote operation, various multimodal feedback methods have been proposed. In fact, these methods are rapidly developing along with the virtual reality industry, which is rapidly increasing in demand. Various attempts, such as exoskeleton, vibration, MR fluid, and ultrasonic wave, are carried out to convey feedback such as force and tactile. Considering maneuverability and installation ability, it is also necessary to use a method suitable for the system. In many cases, the selection of an appropriate number of devices is very important, since the number of feedback devices may preclude natural movement of the human operator. In addition to this reason, research has also been conducted to blend cognitive science while maintaining reality to reduce the number of feedback devices. Therefore, designing robotic control system elements (e.g., motion capture, multimodal feedback method, etc.) must be done beforehand in consideration of the application, destination, and the objective requirements of the system. Since remote control has been actively used, the demand for long distance remote control, including inter-continental
range or earth-space range, is increasing. In this case, to ensure robust remote control performance under the condition that there is data loss in communication and time delay, research is actively being carried out to maintain realism of operators while reducing moving data.

Recently, application of the virtual collaboration system to reduce the amount of data transferred has been actively studied, which generates information locally based on the virtual environment of the remote place pared in advance, not on image information transferred from the remote site. Combinations of the virtual collaboration system and the robotic remote operation system can also create various virtual scenarios and simulate them in advance. Currently, it can also be used for the machine learning method by using various data (e.g., remote environment, situation, operator motion) that are created during remote control. In this case, when the network is unstable, it is possible to interpolate the position or motion data of the operator, which is intermittently inputted to the robot AI so that the remote robot can execute the remotely transferred input commands. It can also be used for operator training and other future possibilities are limitless.

Remote control utilizing the virtual collaboration system is expected to further develop in the future, and it will be able to contribute to making the system more stable through fusion with AI.

Acknowledgement

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT (No. 2017R1C1B5018218, 2017R1C1B5018171). Furthermore, this work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2018R1A5A7023490)

References


