

Article

Development of a High-Speed, Low-Latency Telemanipulated Robot Hand System [†]

Yuji Yamakawa ^{1,*}, Yugo Katsuki ^{2,§}, Yoshihiro Watanabe ³ and Masatoshi Ishikawa ⁴¹ Interfaculty Initiative in Information Studies, The University of Tokyo, Tokyo 153-8505, Japan² Graduate School of Information Science and Technology, The University of Tokyo, Tokyo 113-8656, Japan; yugokatsuki01@gmail.com³ School of Engineering, Tokyo Institute of Technology, Kanagawa 226-8503, Japan; watanabe.y.cl@m.titech.ac.jp⁴ Information Technology Center, The University of Tokyo, Tokyo 113-8656, Japan; ishikawa@ishikawa-vision.org

* Correspondence: y-ymkw@iis.u-tokyo.ac.jp; Tel.: +81-3-5452-6178

[†] This paper is an extended version of our paper published in Katsuki, Y.; Yamakawa, Y.; Watanabe, Y.; Ishikawa, M. Development of Fast-Response Master-Slave System Using High-speed Non-contact 3D Sensing and High-speed Robot Hand. In Proceedings of the 2015 IEEE/RJS International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–2 October 2015; pp. 1236–1241.

‡ Current address: Komaba 4-6-1, Meguro-ku, Tokyo 153-8505, Japan

§ Current address: Sony Imaging Products & Solutions Inc., Tokyo 108-0075, Japan

Abstract: This paper focuses on development of a high-speed, low-latency telemanipulated robot hand system, evaluation of the system, and demonstration of the system. The characteristics of the developed system are the followings: non-contact, high-speed 3D visual sensing of the human hand, intuitive motion mapping between human hands and robot hands, and low-latency, fast responsiveness to human hand motion. Such a high-speed, low-latency telemanipulated robot hand system can be considered to be more effective from the viewpoint of usability. The developed system consists of a high-speed vision system, a high-speed robot hand, and a real-time controller. For the developed system, we propose new methods of 3D sensing, mapping between the human hand and the robot hand, and the robot hand control. We evaluated the performance (latency and responsiveness) of the developed system. As a result, the latency of the developed system is so small that humans cannot recognize the latency. In addition, we conducted experiments of opening/closing motion, object grasping, and moving object grasping as demonstrations. Finally, we confirmed the validity and effectiveness of the developed system and proposed method.

Keywords: telemanipulated system; high-speed vision; high-speed visual feedback; high-speed robot hand



Citation: Yamakawa, Y.; Katsuki, Y.; Watanabe, Y.; Ishikawa, M. Development of a High-Speed, Low-Latency Telemanipulated Robot Hand System. *Robotics* **2021**, *10*, 41. <https://doi.org/10.3390/robotics10010041>

Received: 15 January 2021
Accepted: 26 February 2021
Published: 3 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Recently, there has been significant progress in the development of autonomous robots. Many complex tasks in production processes can now be realized by autonomous robots instead of humans. However, in a dynamically changing environment with unpredictable events, such as the emergence of unknown objects, unexpected changes in the shapes of objects, or irregular motion of objects, human intelligence is required to control the robots appropriately. Thus, research related to Human–Robot Interaction (HRI) has been actively conducted to integrate human intelligence and robot performance. There is a strong demand in HRI for novel technologies and systems that enable remote work in industrial applications.

Therefore, in this research, we focus on a “Telemanipulated system” in the field of HRI [1]. In particular, we pursue the development a high-speed, low-latency telemanipulated robot hand system shown in Figure 1. In this paper, we call the operator of a robot the master and the controlled robot the slave. When it is difficult or unsuitable for humans to

be present at the slave side, or when there are tasks that cannot be carried out by humans alone, a telemanipulated robot system is needed. Examples include handling of dangerous objects (bombs, radioactive materials, etc.) or carrying out tasks that are beyond human capacity (resisting strong physical impacts, high-speed manipulation, etc.).

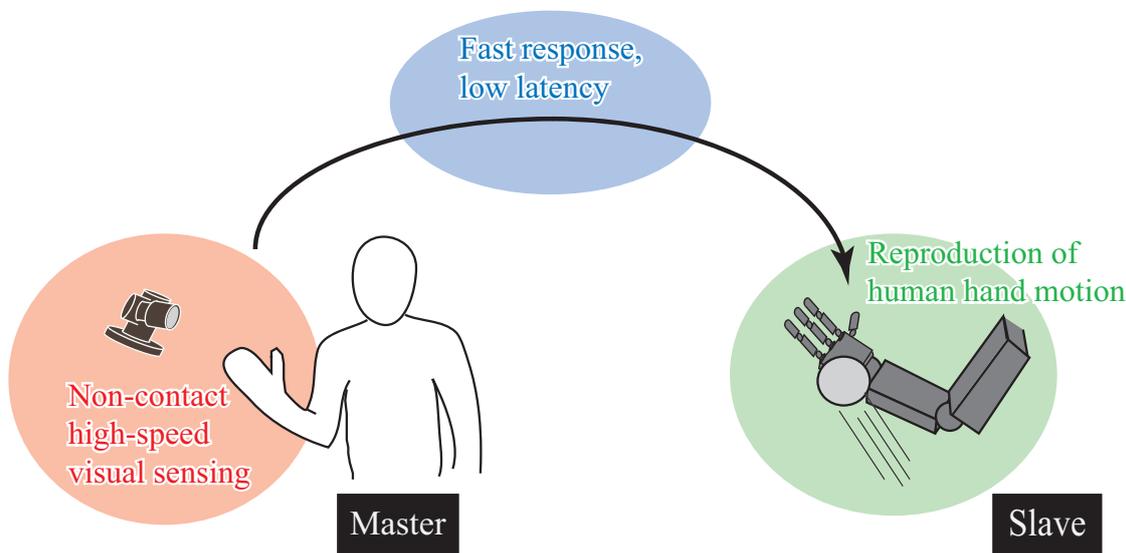


Figure 1. Concept of this research.

2. Related Work

Figure 2 shows the positioning of our research in this field. As the input devices of telemanipulated robot hand systems, three kinds of devices can be considered:

- Joysticks, mice, and GUIs, which are contact-type devices [2]
- Exoskeletal mechanical devices and data gloves, which are contact-type devices
- Vision-based devices, which are non-contact type devices

Conventional telemanipulated robot systems require particular contact-type input devices, such as joysticks, mice, and so on. These devices force humans operators to learn how to use them because their usage is unnatural. Other contact-type input devices include exoskeletal mechanical devices [3], data gloves [4–6], sensors [7–9], or gloves with markers [10]. Some actual telemanipulated systems based on contact-type devices have been successfully employed. However, because it is troublesome for users to wear such devices before operating the system, non-contact-type systems are preferable. Interfaces based on non-contact sensing generally recognize human hand gestures and control a slave robot based on these gestures [11,12]. In the related works in the field of humanoid robotics, the low-cost teleoperated control system for a humanoid robot is developed [13]. In wearable robotics, semantic segmentation has been performed by using Convolutional Neural Networks (CNNs) [14]. Such interfaces are intuitive for users and do not involve the restrictions with contact-type input devices; however, they offer only limited commands for conducting simple tasks, such as adjusting the position or orientation of the slave robot, grasping a static object in front of the robot, and so on. When humans are confronted with tasks that need complex hand motions, it would be difficult to cope with such tasks using these types of interfaces.

In summary, the ideal telemanipulated robot system interface requires an intuitive and natural operating method, and the ability to sense human motion in detail and to reproduce it at the slave side without the constraints associated with contact-type input devices. The positioning of the work described in this paper is shown in Figure 2. A conventional approach that meets such requirements is the “Copycat Hand” developed by Hoshino et al. [15]. They realized accurate and intuitive motion mapping between a

human hand and a robot hand without using contact-type input devices. However, an ideal telemanipulated robot system that is suited to a dynamically changing environment has not yet been developed. The reason for this is that the performance requirements, such as low latency and fast responsiveness, were not satisfied in the conventional telemanipulated system. It is known that high-latency and low responsiveness degrade the performance of the telemanipulated system [16]. The fundamental factors caused high latency and slow responsiveness are considered to be the low frame rate of cameras, low-speed image processing, and the slow response of robots. In the work described in this paper, we developed an entirely new telemanipulated system to overcome the problems of high latency and slow responsiveness.

Based on the above discussion, we propose a fast-response telemanipulated robot hand system using high-speed non-contact 3D sensing and a high-speed robot hand [17,18]. An operator can adapt to dynamic changes in the environment at the slave side thanks to the high-speed sensing. Very fast human hand motion can be measured without much loss of information between each frame thanks to the high-speed sensing. Furthermore, our proposed system realizes high responsiveness, and the latency from input to output is so low that humans cannot recognize it.

The system structure is described in detail in Section 3. Proposed methods for 3D visual sensing, motion mapping between the human hand and the robot hand, and the robot hand control are explained in Section 4. We confirmed the effectiveness of our proposed system through three types of experiments (basic motion, ball catching, and falling stick catching), and the results are discussed in Section 5. Finally, we summarize the conclusions obtained in this study in Section 6. This paper extends the detail and analysis of the proposed method and the additional simulation and experiments from the previous paper [18]. In addition, the contributions of this paper are summarized as follows:

- Detailed explanation of high-speed 3D visual sensing, motion mapping and robot hand control
- Proposal of general mapping algorithm method
- Simulation analysis for evaluating our high-speed, low-latency telemanipulated system
- Static and dynamic object catching to show effectiveness of system
- Analysis of dynamic object catching (falling stick catching)

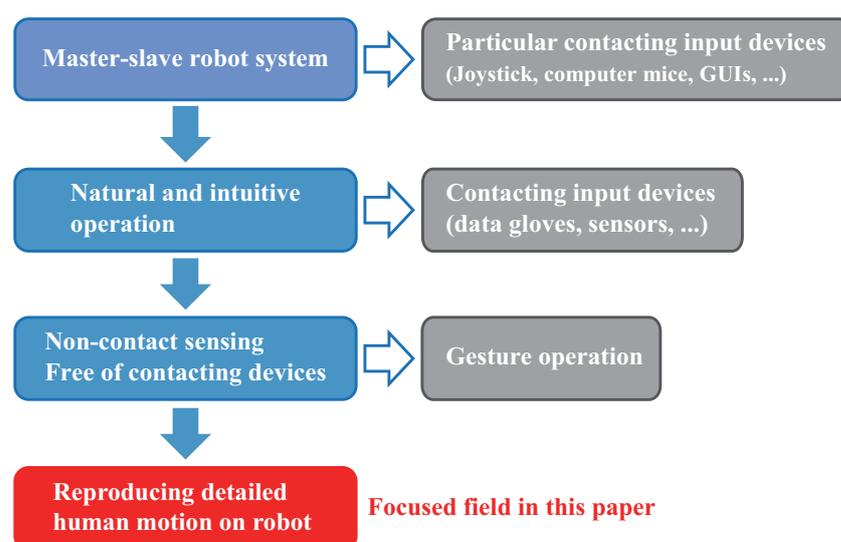


Figure 2. Positioning of our work [18] (© 2015 IEEE).

3. High-Speed Telemanipulated Robot Hand System

A conceptual diagram of the overall system is shown in Figure 3. Our proposed system is composed of:

1. A high-speed vision system (Section 3.1)
2. A high-speed robot hand (Section 3.2)
3. A real-time controller (Section 3.3)

The high-speed vision system captures human hand images at the master side, and essential information about the human hand motion is extracted by the image processing PC. The information is transmitted to the real-time controller at the slave side via Ethernet. Then, the high-speed robot hand system is controlled by the real-time controller based on the information received from the master side. A detailed explanation of each component of our proposed system is given below.

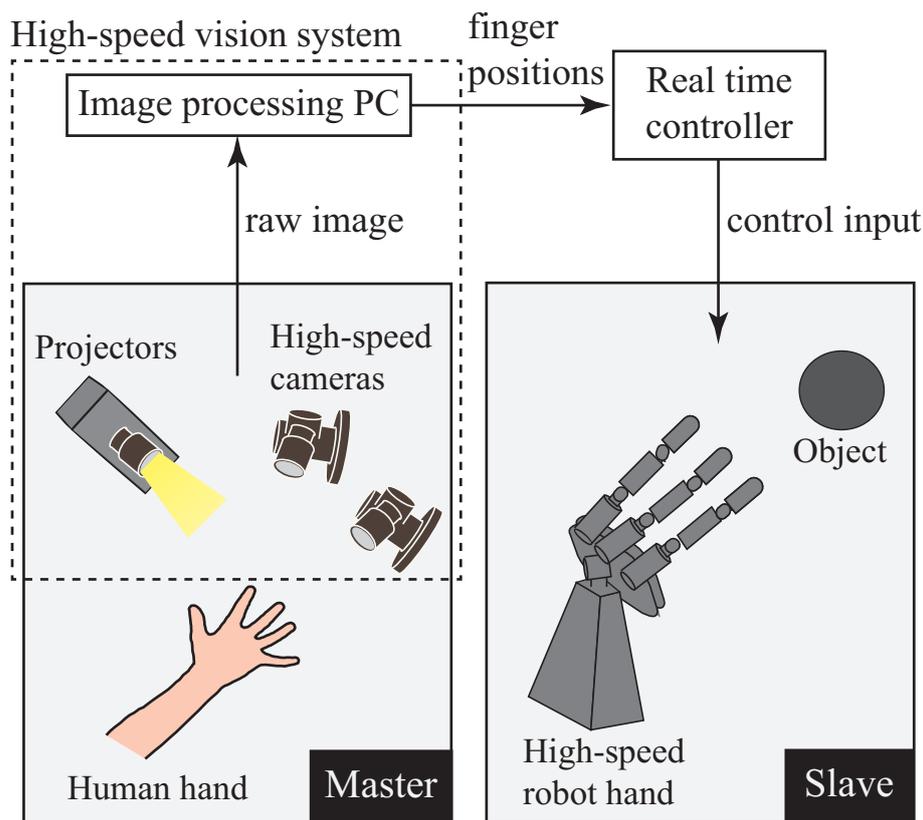


Figure 3. Conceptual illustration of the high-speed, low-latency telemanipulated robot hand system [18] (© 2015 IEEE).

3.1. High-Speed Vision System

The high-speed vision system shown in Figure 4 is composed of two high-speed cameras, one high-speed projector, and an image-processing PC. IDP Express R2000 is used as the high-speed camera. It is small and lightweight (35 mm × 35 mm × 34 mm, 90 g), and it has 250,000 pixels. It has a frame rate of up to 2000 frame-per-second (fps), and two high-speed cameras can be used simultaneously with one acquisition board. Captured images are transmitted to the image-processing PC in real time. The performance of the image-processing PC is as follows: CPU: Intel® Xeon® E5640 (double processors 2.67 GHz and 2.66 GHz) (Intel Corporation, Santa Clara, CA, USA); Memory: 96 GB; and Operating System: Windows 7 64 bit. The image-processing program was implemented and run in Microsoft Visual Studio. A Digital Light Processing (DLP) LightCrafter 4500 is also used as a high-speed projector. This projector uses a Digital Mirror Device (DMD), which can project images at high speed. A single-line pattern is projected on the operator's hand from the projector. The two high-speed cameras and the high-speed projector are synchronized at 500 fps, and the relative positions of the two high-speed cameras are precisely known via calibration carried out in advance. Details of the image processing are given in Section 4.1.

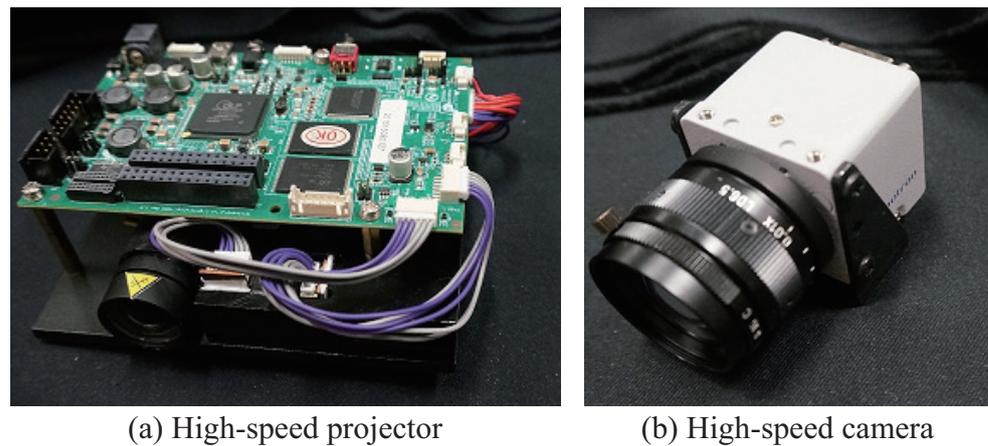


Figure 4. High-speed vision system [18] (© 2015 IEEE).

The setup of the sensing system is shown in Figure 5. Images were captured from above, and so the target human hand should be placed under this sensing system. The distance between the human hand and the sensing system was about 30 cm in experiments.

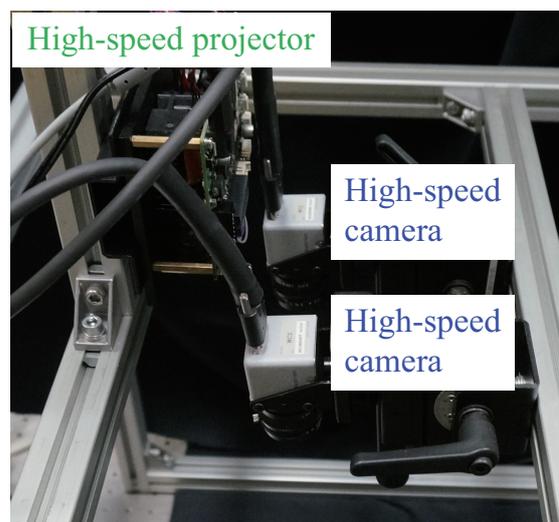


Figure 5. Setup of sensing system [18] (© 2015 IEEE).

3.2. High-Speed Robot Hand

The high-speed robot hand (Figure 6) [19] has three fingers: an index finger, a left thumb, and a right thumb. Each finger is composed of a top link and a root link. The index finger has two degrees of freedom (2-DOF), and the other fingers (left and right thumbs) have 3-DOF. Furthermore, each hand has two wrist joints. The wrist part employs a differential rotation mechanism and moves about two axes: a bending-extension axis and a rotation axis. The high-speed robot hand can be controlled at a sampling time of 1 ms by the real-time control system. The joint angles of the hand can be controlled within 1 ms by a Proportional-Differential (PD) control system. The details of the robot hand control are given in Section 4.3.

The joints of the hands can be closed at a speed of 180 deg/0.1 s, which has been considered to be a sufficiently high level of performance to reproduce fast human hand motion. We conducted an experiment to confirm whether this is true, and the results are explained in Section 5.

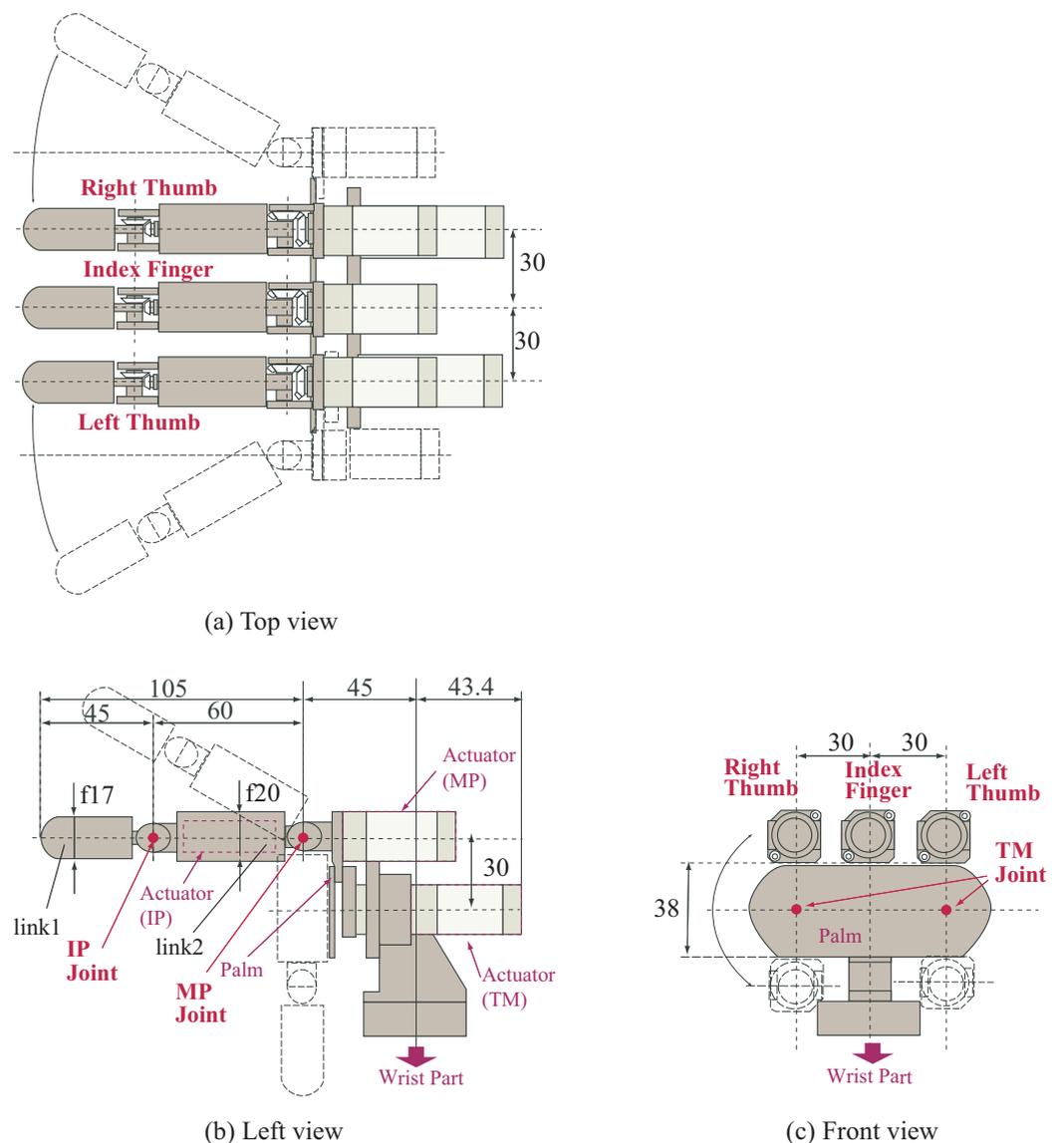


Figure 6. Mechanism of high-speed robot hand [19] (© 2004 IEEE).

3.3. Real-Time Controller

To perform high-speed and low-latency processing such as Ethernet communication and calculations of robot hand trajectory and joint angle control, we need a real-time controller that can run with a sampling time of 1 ms. Thus, we used a commercial real-time controller produced by dSPACE. By using this real-time controller, we can achieve robot hand control with visual feedback at 1 kHz. The real-time controller has counter, digital/analog output, and Ethernet communication function.

4. Strategy for Telemanipulated Robot Hand System

In this section, the proposed method for realizing fast-response telemanipulation is explained. The strategy can be divided into three components:

1. 3D Hand Position Sensing (Section 4.1)
2. Motion Mapping to Robot Hand (Section 4.2)
3. Joint Angle Control of Robot Hand (Section 4.3)

First, finger motion data are obtained using the high-speed vision system. Then, structural differences between the human hand and the robot hand have to be adjusted in order to make the operation natural and intuitive. The adjusted position data of the fingers of the human hand need to be converted to joint angle data for controlling the high-speed

robot hand based on inverse kinematics. Finally, the joint angles of the high-speed robot hand are controlled to follow the joint angle data, which is updated every 2 ms. The flow of the proposed method is shown in Figure 7, and the details of our proposed method are given below.

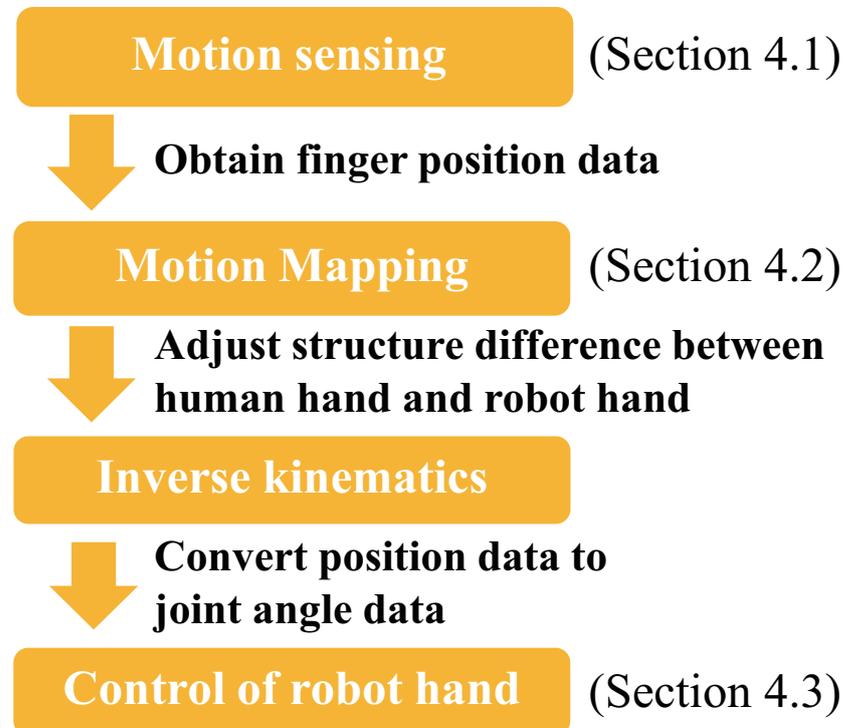


Figure 7. Flow of proposed method [18]; Motion sensing, Motion mapping and Control of robot hand are explained in Section 4.1, Section 4.2 and Section 4.3, respectively. (© 2015 IEEE).

4.1. High-Speed Visual 3D Sensing of Human Hand

The role of our sensing system is to obtain three-dimensional (3D) position data of the fingers of the human hand. Since non-contact sensing using a vision system is employed, there is no need for any contact-type devices, sensors, or markers on the human hand. However, the fingers of the human hand do not have sufficient features, and a passive sensing method is not appropriate. Therefore, we selected an active sensing method to obtain more-precise 3D data of the finger positions. We assume that the target human hand is moving while its 3D data are being sensed, and it is desirable to select a one-shot sensing method to avoid loss of information between each frame. Furthermore, an ideal sensing system needs to be capable of high-speed real-time processing to adapt to dynamic changes in the environment. Based on these requirements, we selected an active sensing method using a line pattern. The detailed process of the high-speed non-contact 3D sensing of the human hand is explained below. A conceptual diagram is shown in Figure 8, and the procedure of the visual sensing is as follows:

1. Binarization of captured images (Section 4.1.1)
2. Finding the corresponding points (Section 4.1.2)
3. 3D information calculation of corresponding points (Section 4.1.3)
4. Calculation of centroid of each finger position (Section 4.1.4)

The following explains the detailed process of the high-speed non-contact 3D sensing of the human hand.

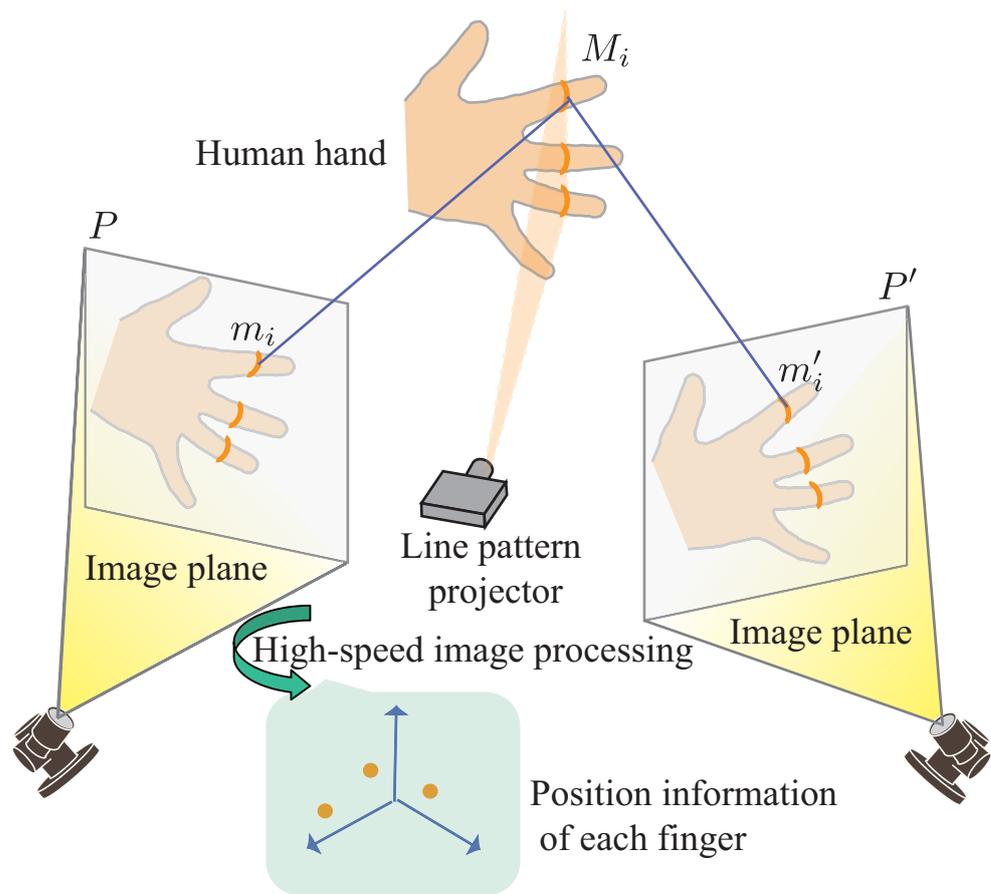


Figure 8. High-speed non-contact sensing [18] (© 2015 IEEE).

4.1.1. Binarization of Captured Images

First, two images captured simultaneously by the two high-speed cameras are binarized. A captured camera image is represented by $f(x, y)$, the binarization threshold as θ , and the binarized image as $g(x, y)$. The following equation represents the binarization process, in which the threshold should be selected based on the environment at the master side:

$$g(x, y) = \begin{cases} 1 & (f(x, y) \geq \theta) \\ 0 & (f(x, y) < \theta). \end{cases} \quad (1)$$

By performing this binarization processing, the parts projected by the projector are extracted. As a result, the image processing explained below becomes easier.

4.1.2. Finding Corresponding Points

This section explains how to find corresponding points on both images of the left and right cameras. Figure 9 shows the illustration of finding corresponding points. This procedure can be executed into two steps as follows:

1. Scan the binarized image of the left camera in the X direction while scanning the binarized image of the same image line by line in Y direction.
2. When a point, which is shown as the red square in Figure 9, is found, the epipolar line on the binarized image of the right camera is scanned, and the found point is used as the corresponding point.

By using the proposed method, we can find the corresponding points on both images at high speed.

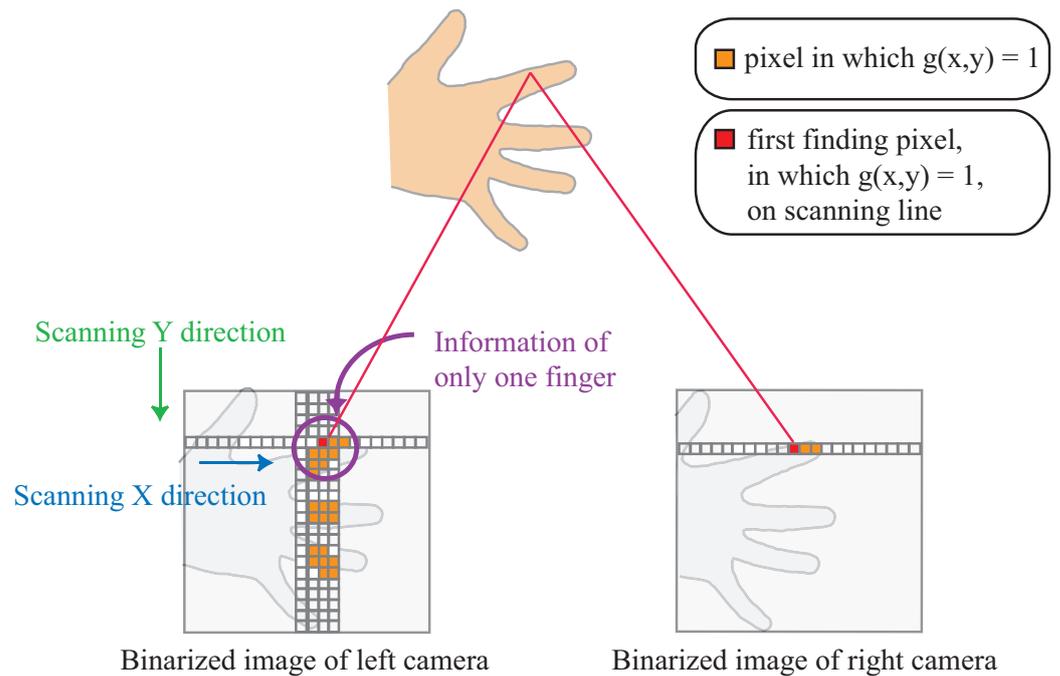


Figure 9. Finding corresponding points.

4.1.3. 3D Information of Corresponding Points

The pixels that correspond to the finger part on which the line pattern is projected have a value of 1 after binarization, whereas the other parts have a value of 0. The 3D information of each finger can be obtained based on this feature.

The 3D points to be measured are represented by $M_i(x, y, z)(i = 1, \dots, N_p)$, where N_p is the number of 3D points measured in one cycle. In this work, we set $N_p = 3$, because we need to obtain the positions of the three fingers of the human hand. The coordinates of each pair of corresponding points are defined as $m_i = [u; v; 1]^T$ and $m'_i = [u'; v'; 1]^T$. The projection matrices of the high-speed cameras are P and P' , which include intrinsic and extrinsic parameters. s and s' are scalar values. The constraint equations are:

$$sm_i = PM_i, s'm'_i = P'M_i. \tag{2}$$

Here, P and P' are parameters that are known in advance by using camera calibration, and m_i and m'_i are measured data obtained by the high-speed camera and image processing. Equation (2) represents the correspondence relationship between m_i and m'_i and M_i based on the projection matrices P and P' . These constraint equations mean that M_i is located on the intersection of two lines of sight from the corresponding points m_i and m'_i . M_i can be obtained by taking advantage of this relation.

Combining the two equations expressed in Equation (2) as one equation yields

$$BM = b, \tag{3}$$

where

$$B = \begin{bmatrix} up_{31} - p_{11} & up_{32} - p_{12} & up_{33} - p_{13} \\ vp_{31} - p_{21} & vp_{32} - p_{22} & vp_{33} - p_{23} \\ u'p'_{31} - p'_{11} & u'p'_{32} - p'_{12} & u'p'_{33} - p'_{13} \\ v'p'_{31} - p'_{21} & v'p'_{32} - p'_{22} & v'p'_{33} - p'_{23} \end{bmatrix}, \tag{4}$$

$$b = [p_{14} - up_{34} \quad p_{24} - vp_{34} \quad p'_{14} - u'p'_{34} \quad p'_{24} - v'p'_{34}]^T. \tag{5}$$

Here, p_{ij} and p'_{ij} are the values of the component (i, j) of the matrices $(P$ and $P')$, respectively.

Thus, the 3D information of the target point, which corresponds to the finger part in this work, can be calculated by

$$M = B^+b, \quad (6)$$

where B^+ is the pseudo inverse matrix of B .

Then, the same procedure is repeated while scanning the pixel cloud from the left to the right in the next row of the left image.

4.1.4. Calculation of Centroid of Each Finger Position

Last, the position of each finger is determined. We assume that the number of fingers on which a line pattern is projected is three (thumb, index finger, and middle finger). This number corresponds to the number of fingers on the high-speed robot hand. The obtained 3D points are clustered in accordance with the separate position of each finger. Each set of clustered 3D points is classified and defined as each finger position:

$$X_j = \frac{1}{N_j} \sum_{i=\alpha_j}^{N_j+\alpha_j-1} M_i(x, y, z), (j = 1, 2, 3). \quad (7)$$

Here, j is an index identifying each finger, N_j is the number of 3D points clustered on each finger, X_j is the 3D centroid coordinates of one finger, and α_j is a starting number of 3D points clustered on each finger.

This cycle of the process is repeated every 2 ms in real time for each frame captured by the high-speed cameras.

4.2. Motion Mapping to Robot Hand

The structure of the human hand and that of a robot hand are often different, except for robot hands that imitate the structure of the human hand. It is highly likely that the method of mapping human hand motion to robot hand motion will have a large influence on the performance of the telemanipulated robot system. Therefore, motion mapping is an important factor, and some studies related to motion mapping have been conducted [20,21].

4.2.1. Design Principle

The fundamental principle of motion mapping is to reproduce human motion on the slave side robot in a natural and intuitive way for the operator. Therefore, we think that the factors that should be considered are the following:

- Positional relationship of fingers
- Direction and amount of motion
- Movable range of fingers

First, the positional relationship of the fingers should be the same between the human hand and the robot hand. For example, when the motion of the fingers on the right side of the human hand is mapped to the fingers on the left side of the robot hand, the operator will inevitably feel a sense of disorientation during operation, and it will take time for the operator to get used to this unnatural motion reproduction on the robot.

Furthermore, the direction of motion of the fingers on the human hand needs to be the same as that of the fingers on the robot hand. In addition, the amount of motion of the human hand needs to be adjusted while considering the structural differences between the human hand and the robot hand.

Lastly, the movable range of the fingers needs to be properly decided because it is generally different between the human hand and the robot hand.

4.2.2. Mapping Algorithm

This section explains a mapping algorithm from the human hand to the robot hand based on the obtained the information of the human hand. Firstly, we propose a general

mapping algorithm. After that, we also describe a detailed calculation of the proposed mapping algorithm.

Proposal of General Mapping Algorithm

We describe a general formulation of mapping algorithm from the human hand to the robot hand. We define the position vectors of the fingertips on the master (the human hand) and the slave (the robot hand) sides as X_{Ma} and X_{Sa} ($a = 1, 2, 3, \dots$), respectively. Then, we represent internally dividing points on the master and the slave sides as follows:

$$X_{MG} = \sum_{n=1}^a X_{Mn} m_n \quad (8)$$

$$\sum_{n=1}^a m_n = 1 \quad (9)$$

$$X_{SG} = \sum_{n=1}^a X_{Sn} s_n \quad (10)$$

$$\sum_{n=1}^a s_n = 1 \quad (11)$$

Assuming that the centroid the fingertips of the robot hand can be expressed by X_{SG} , we can calculate each fingertip position of the robot hand by using the rotation matrix R and the gain matrix G as follows:

$$X_{Sa} = X_{SG} + GR(X_{Ma} - X_{MG}) \quad (12)$$

In this work, we set $a = 3$, $m_1 = m_2 = m_3 = s_1 = s_2 = s_3 = \frac{1}{3}$ because the robot hand has three fingers. Moreover, in the actual experiments explained in Section 5, the high-speed vision system recognizes the positions of three fingers of the human hand and the centroid of each finger position, as described in Section 4.1. Next we explain the concrete calculation of this proposed method.

Concrete Calculation of Proposed Mapping Algorithm

The initial posture and correspondence relationship of each finger is shown in Figure 10. X_{M1} , X_{M2} , and X_{M3} are the position coordinates of the fingers on the human hand (X_{M1} : thumb; X_{M2} : index finger; and X_{M3} : middle finger). The centroid of those coordinates is defined as X_{MG} , which can be calculated by

$$X_{MG} = \frac{X_{M1} + X_{M2} + X_{M3}}{3} \quad (13)$$

Here, X_{S1} , X_{S2} , and X_{S3} are the position coordinates of the fingers on the robot hand (X_{S1} : right thumb; X_{S2} : index finger; and X_{S3} : left thumb). The centroid of those coordinates is also defined as X_{SG} , which can be obtained in the same way as the calculation of X_{MG} . Initial offset values are added to each of X_{M1} , X_{M2} , and X_{M3} in order to correct size differences between the human hand and the robot hand.

Then, the following equation is applied to adjust for the difference in motion between the human hand and the robot hand:

$$X_{Sa} = X_{SG} + A(X_{Ma} - X_{MG}) \quad (a = 1, 2, 3). \quad (14)$$

The value of variable A should be decided based on the size of the human hand, which differs from person to person; this time, we set $A = 1.2$. After X_{S1} , X_{S2} , and X_{S3} are calculated, those position coordinates are converted to joint angle information through the inverse kinematics of the robot hand. As a result, the desired joint angles of the fingers of the robot hand for duplicating the motion of the human hand can be obtained.

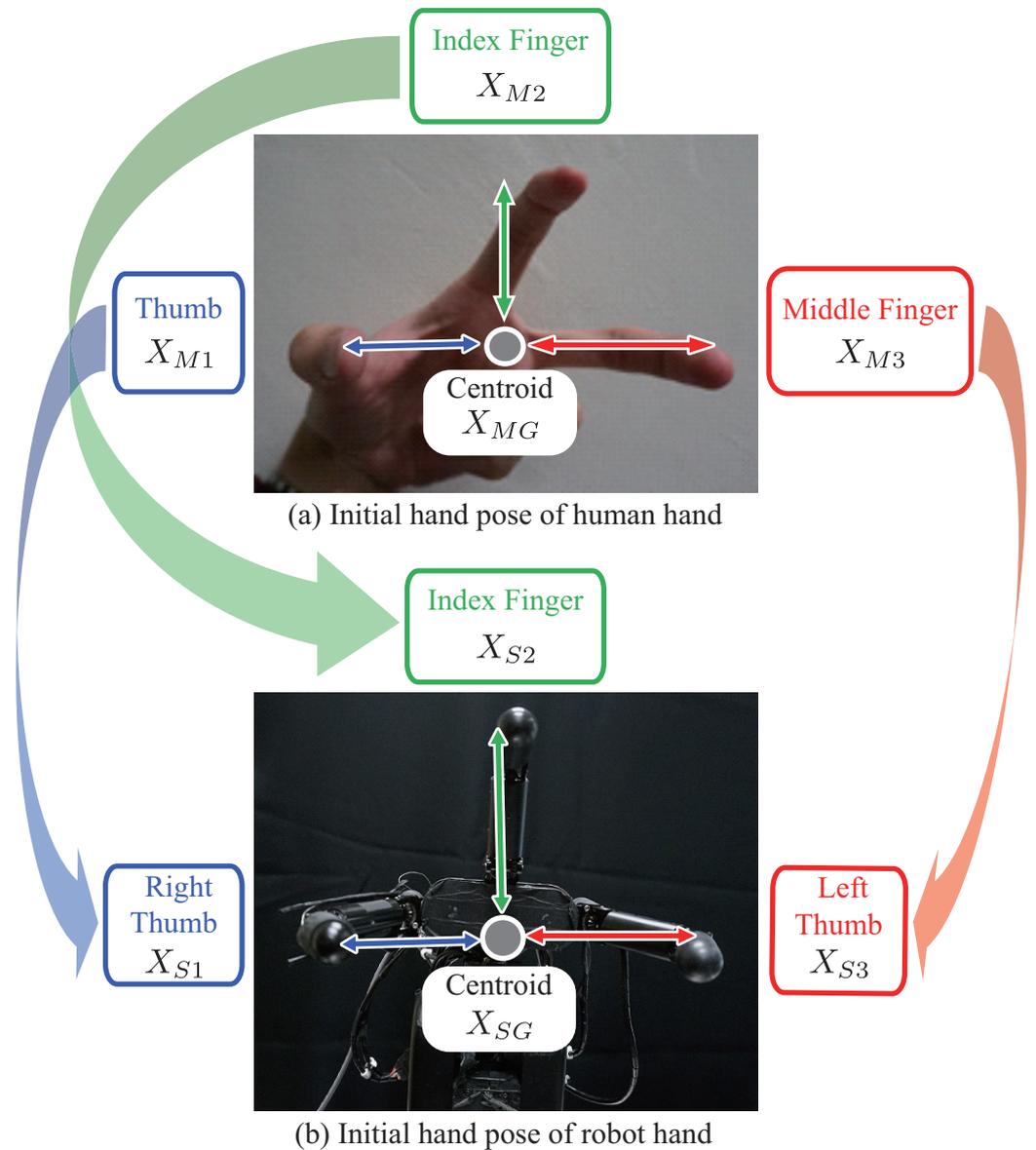


Figure 10. Initial posture and correspondence relationship of each finger [18] (© 2015 IEEE).

4.3. Joint Angle Control of Robot Hand

To follow the desired joint angles of the robot hand, the joint angles of the robot hand are controlled by a Proportional-Differential (PD) control law. Namely, the following torque input τ described by Equation (15) is applied to the each actuator installed in the robot hand:

$$\tau = k_p(\theta_r - \theta) + k_d(\dot{\theta}_r - \dot{\theta}), \tag{15}$$

where θ_r is the desired joint angle of the robot hand, θ is the actual joint angle of the robot hand and can be measured by an optical encoder installed in the servo motor, and k_p and k_d are the proportional and differential parameters, respectively.

4.4. Advantages and Limitations of Proposed Method

The characteristics and advantages of the developed telemanipulated robot hand system have three aspects: (1) non-contact, high-speed 3D visual sensing of the human hand; (2) intuitive motion mapping between human hands and robot hands; and (3) low-latency, fast responsiveness to human hand motion. Consequently, our system possesses the

performance of high speed beyond human and low latency that human cannot recognize. The system performance is evaluated in the next section.

On the other hand, the limitations of the developed telemanipulated robot hand system are the following: (1) the fingertips need to be brightened with a projector; and (2) the human hand has to be placed in a fixed place for that purpose. These limitations will be solved by using advanced image processing and hand tracking [22,23] technologies in the future.

Finally, we compare our method with other methods such as using RGB-Depth camera, event camera, and machine learning technology. If we use RGB-Depth camera or machine learning technology, the frame rate becomes slow or the processing time becomes long. As a result, the latency may become long, and high-speed, low-latency telemanipulated operation cannot be achieved. In addition, if we use event camera, visual sensing may become more difficult when the movement of the human hand is extremely slow or stopped. Thus, we can consider that it is not suitable to use the other methods in the telemanipulated system, and the proposed method is better than the other methods.

5. Experiments

In this section, we describe experiments conducted using our proposed system and method. First, we confirmed the validity of the proposed method through simulation (Section 5.1). Second, we investigated the maximum motion speed of a finger on the human hand and confirmed that our telemanipulated robot system could reproduce its motion (Section 5.2). In addition, the latency from input to output was examined, and we evaluated the response speed. In addition, we realized two demonstrations: “Ball catching” (Section 5.3) and “Falling stick catching” (Section 5.4). A video of the demonstration is available on our website [24].

5.1. Simulation

We verified the validity of the proposed motion sensing and mapping method between the human hand motion and the robot hand motion by simulation. Figure 11 shows the initial state of the human hand and the robot hand in simulation environment. In Figure 11, the left and right panels show the initial states of the human hand and the robot hand in simulation, respectively. The robot hand structure is based on the actual high-speed robot hand described in Section 3.2.

Figure 12 illustrates the simulation result (trajectories of human fingertip and robot hand fingertip) when the human hand fingers close from the initial state. Figure 12a,b shows trajectories of human fingertip and robot hand fingertip, respectively. From this result, we found that the trajectory of the human fingertip and the trajectory of the fingertip of the robot hand are similar, and mapping is performed successfully according to human intuition.

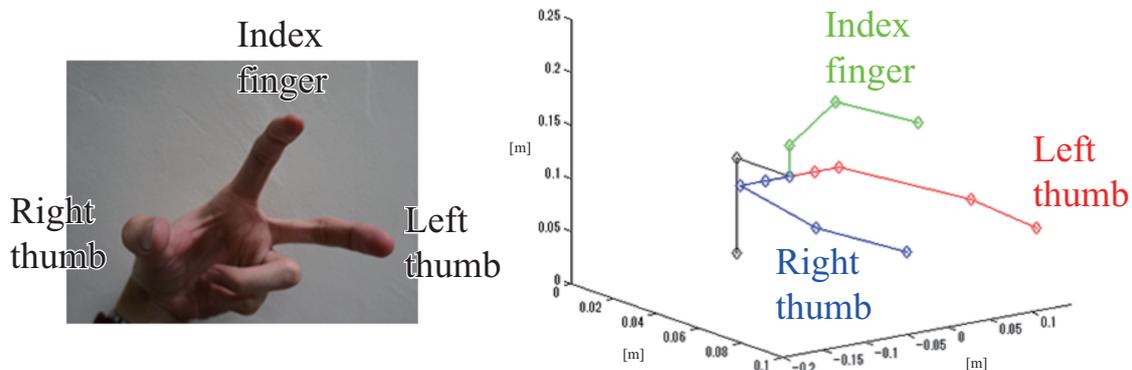


Figure 11. Initial state of simulation.

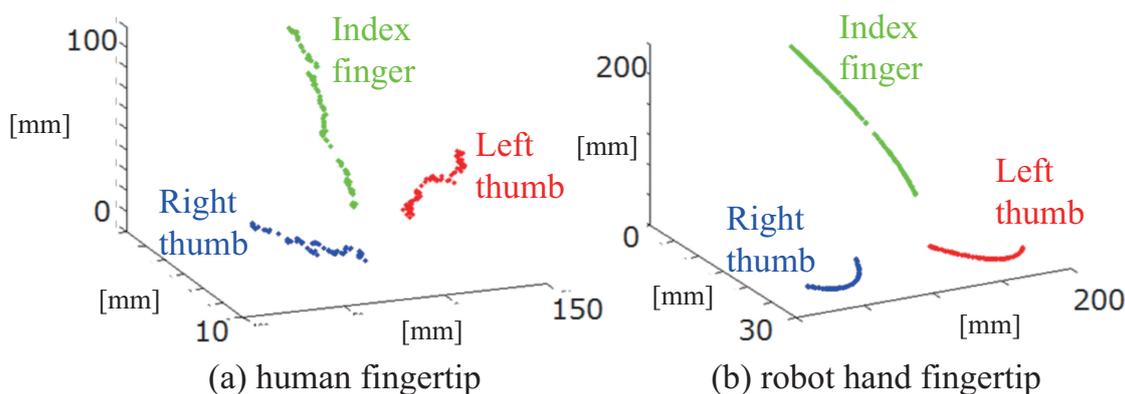


Figure 12. Mapping result.

5.2. Evaluation of Motion Synchronization

The high-speed telemanipulated robot hand system should have sufficient performance to reproduce high-speed motion of the fingers on the human hand, as described in Section 3. From the viewpoint of motion synchronization between human hand motion and robot hand motion, we conducted experiments to evaluate the following elements:

1. High-speed motion of human finger (Section 5.2.1)
2. Reproduction of high-speed motion (Section 5.2.2)
3. Fast responsiveness (Section 5.2.3)
4. Opening and closing motion as a basic demonstration (Section 5.2.4)

5.2.1. High-Speed Motion of Human Finger

We measured the high-speed motion of a human hand finger using a 1000 fps high-speed camera. The second joint of the thumb performed reciprocating motion as fast as possible, and the amplitude of the joint angle was $1/(6\pi)$. Figure 13 shows the experimental result of the high-speed motion of the human finger. In this figure, the blue line represents normalized time-series data of the position of the human finger. Reciprocating motion was performed about five times per second. From this result, we could measure the high-speed motion of the human finger using the high-speed vision system and our proposed visual 3D sensing method. Moreover, we found that our developed system can reproduce the high-speed motion of the human hand, because the performance of the high-speed robot hand is beyond that of the human hand.

In the next section, we describe experimental results of reproducing high-speed motion with our developed system.

5.2.2. Reproducing High-Speed Motion

The same motion as that of the human finger was reproduced on the robot hand with our developed system. The results of this experiment are shown in Figure 13. The dotted green line represents normalized time-series data of the joint angle of a robot finger that corresponds to the human finger. The two data are aligned on the same time series. The values of the dotted green line follow the changes in the values of the blue line with an extremely small latency, as described above. This latency is evaluated in more detail in the next section. From these results, it can be found that the fast motion of the human finger was successfully reproduced on the high-speed robot at the slave side using our proposed telemanipulated system and method.

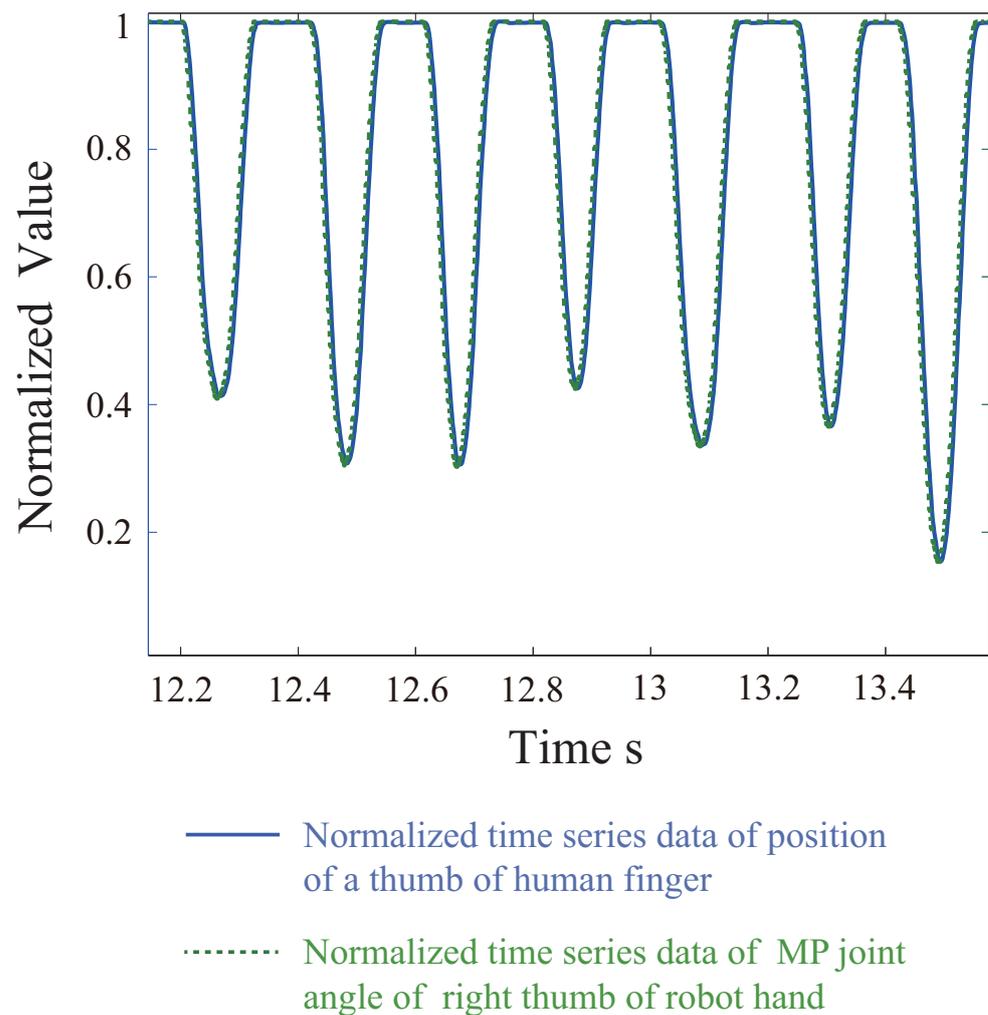


Figure 13. Reproduction of high-speed motion of a human finger [18] (© 2015 IEEE).

5.2.3. Fast Responsiveness

The latency of our proposed system was evaluated based on the graph in Figure 14. The blue line indicates the reference joint angle values, and the green line indicates the actual joint angle values. The command value of the reference joint angle was input every 2 ms. This means that real-time image processing was conducted at 500 fps on the master-side sensing system. In addition, the time for the joint angle to achieve the reference joint angle was about 15 ms. Moreover, the latency from data (result of image processing) sending to control signal (torque input to servo motor) output was added [25].

Based on these, the total latency of the developed system can be derived by adding the image processing time (2 ms), the data communication time (3 ms [25]), and the response time of the robot hand (15 ms) and calculated by Equation (16). As a result, we calculated the latency from an input acquired by the high-speed sensing system to the output at the high-speed robot hand. The result is indicated in Equation (16) [25]:

$$\begin{aligned}
 & 2 \text{ ms (image processing)} \\
 & + 3 \text{ ms (real time control from data reception to control signal output)} \\
 & + 15 \text{ ms (response latency of high-speed robot hand)} \\
 & \leq 20 \text{ ms.}
 \end{aligned} \tag{16}$$

The total latency was estimated to be less than 20 ms. We evaluate this latency in more detail below.

Grill-Spector et al. studied the visual recognition speed of the human eye [26]. The differences in success rates of detection, categorization, and identification of an object in an image were investigated for different exposure durations. The success rate was found to be better when the exposure duration of the image was 33 ms than when it was 17 ms. Based on this, we assume that it is difficult for humans to detect a latency of less than 30 ms. Actually, we developed a system called “Rock-Paper-Scissors Robot System with 100% Winning Rate” [27], which has the same latency as our proposed telemanipulated robot system. In that system, the latency from input to output is so low that humans cannot recognize it. A video of this system is available on our website [28]. Therefore, we consider that humans cannot recognize the latency of our proposed telemanipulated robot system either.

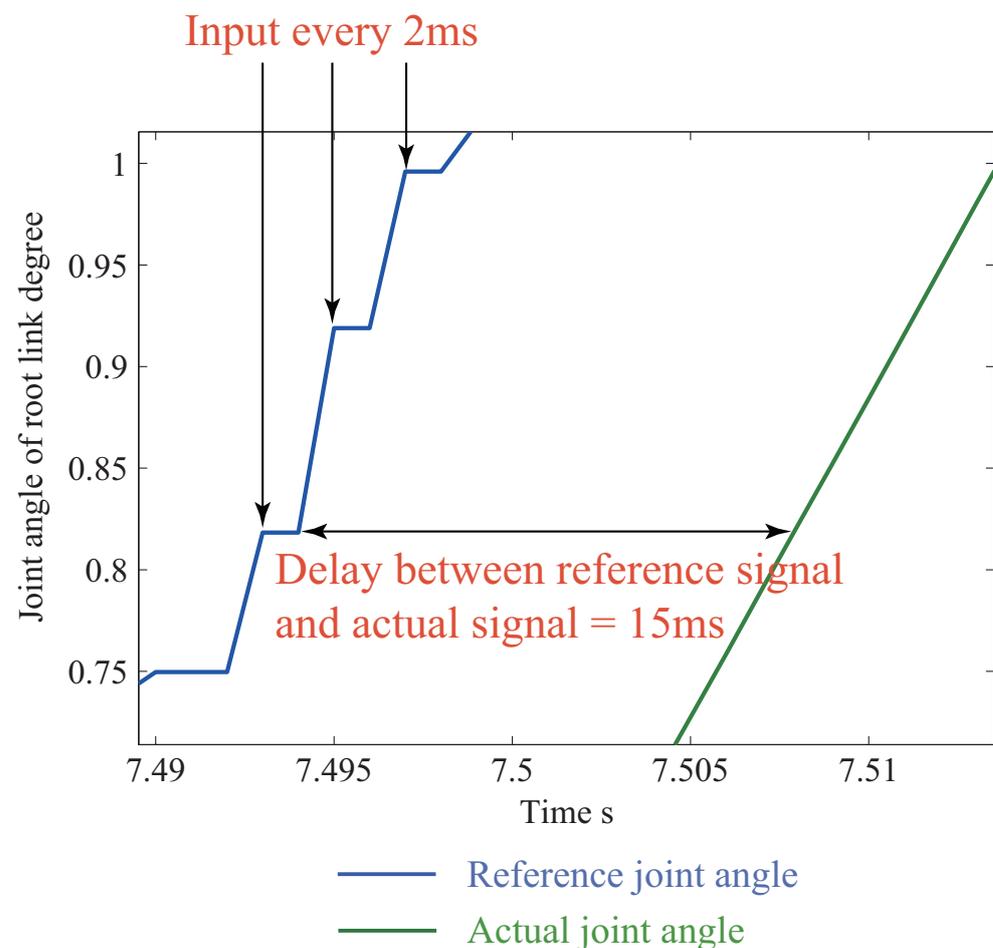


Figure 14. Latency for actual joint angle to achieve reference joint angle [18] (© 2015 IEEE).

5.2.4. Opening and Closing Motion

Here, we show continuous photographs of the experimental results of opening and closing motion, which can be considered to be one basic demonstration of the telemanipulated system, in Figures 15 and 16. The experimental data is shown in Figures 13 and 14. It can be seen in these figures that the motion synchronization between the human hand and the robot hand can be achieved successfully.

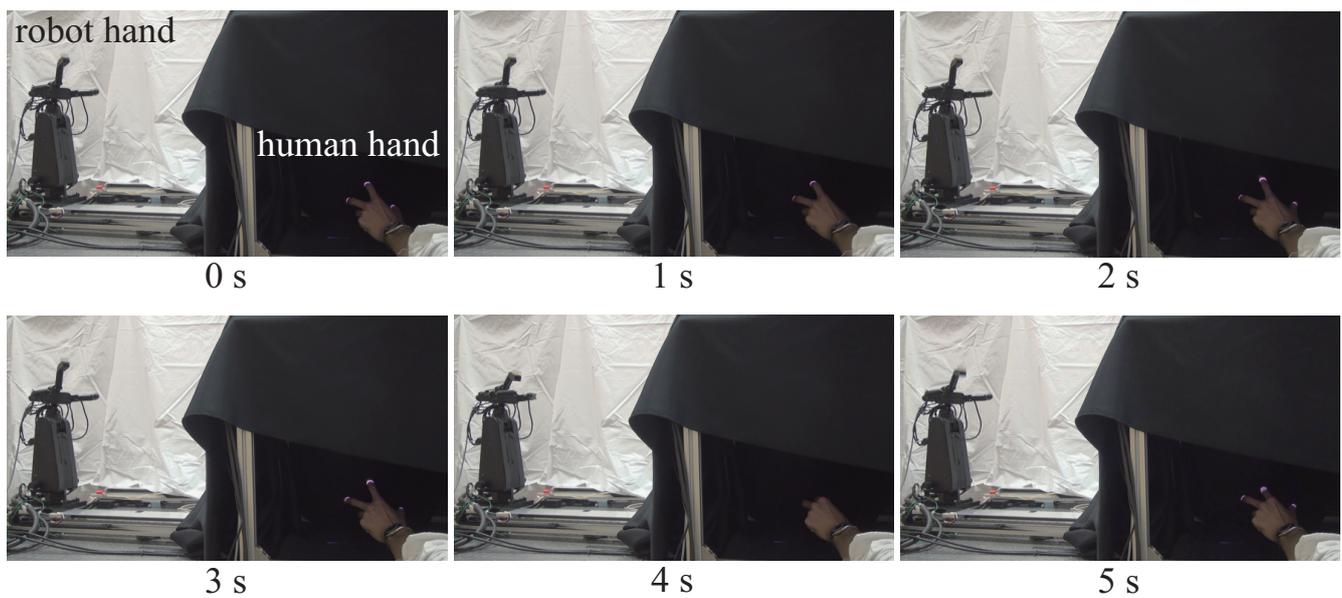


Figure 15. Experimental result of hand opening and closing motion.

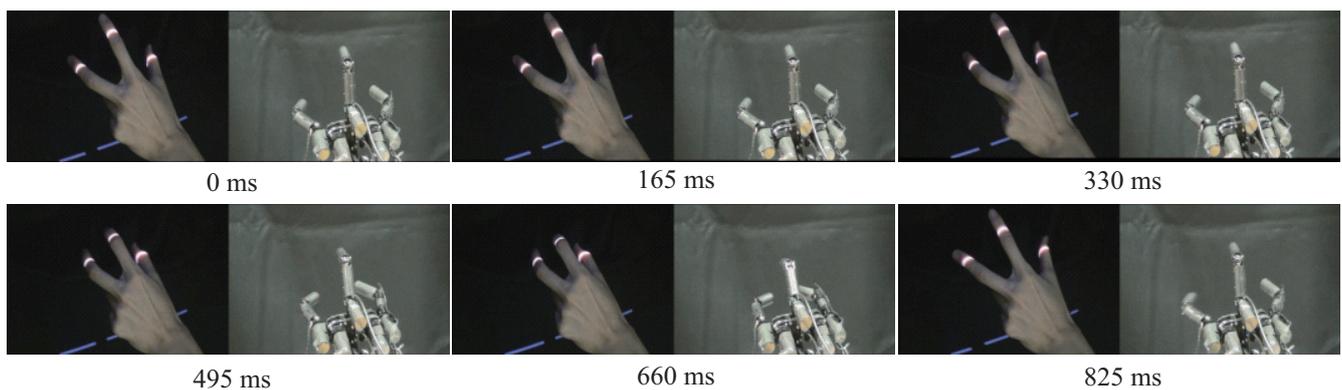


Figure 16. Experimental result of hand opening and closing motion (enlarged) [18] (© 2015 IEEE).

5.3. Ball Catching

Figure 17 shows an experimental result of ball catching, which can be considered to be one demonstration of static object grasping. In this experiment, a human subject on the slave side presented the ball in front of the robot hand at random timing, and a human subject on the master side grasped the ball by master–slave manipulation. From the experimental result, it was found that the proposed system was able to flexibly respond to changes in the slave environment and perform tasks at the discretion of the human subjects.

5.4. Falling Stick Catching

Furthermore, to confirm the effectiveness of our proposed system in a task that needs high-speed performance, we conducted an experiment to catch a falling object with the robot hand at the slave side, which can be considered to be one demonstration of dynamic object grasping. A stick with a length of 16 cm, which was made of polystyrene, was used as the target object. The stick was dropped from a height of 40 or 70 cm above the robot hand at random timing. Trials were conducted in two master–slave systems. One was the proposed system, and the other was the same system but with the only difference being that it used a 30 fps vision system. Twenty trials were conducted on each system.

It is said that the reflexes of a human in a simple task are about 220 ms [29], which is shorter than the time required for the stick to fall 40 cm, so, even if it is dropped from 40 cm above, the human can catch it with the robot hand. From the drop point on the slave

side, a human drops the stick at a random timing in each trial. The operator on the master side operates the robot hand and tries to catch the stick. The distance between the human and the high-speed robot hand was about 90 cm, and the height of the human eye was almost the same as the height of the robot hand.

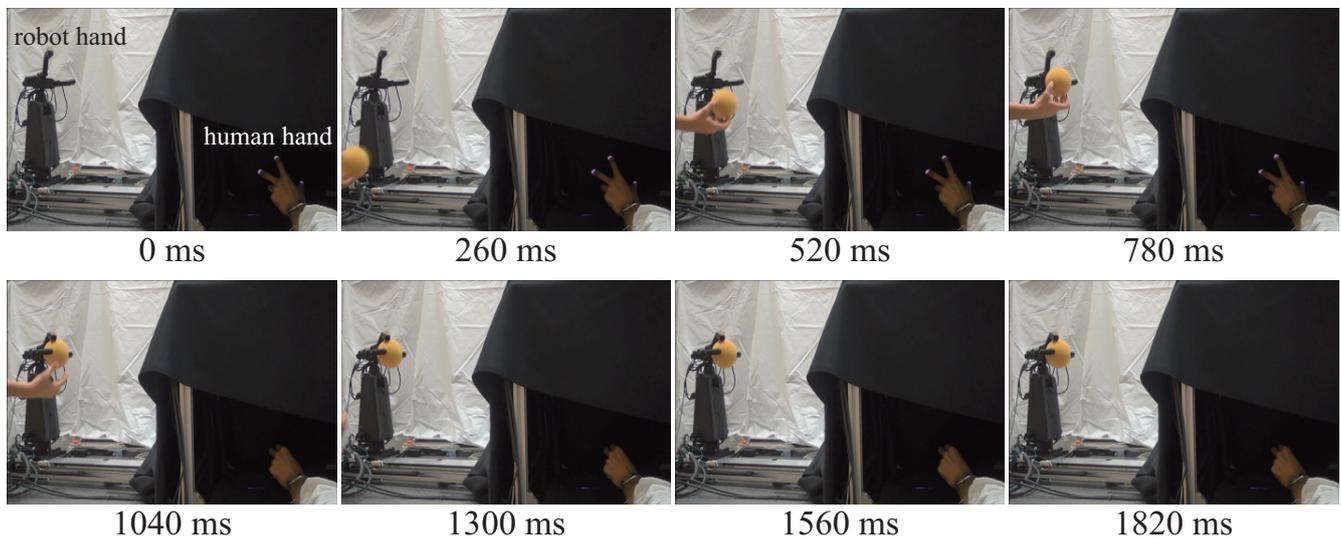


Figure 17. Experimental result of ball catching.

Figure 18 shows the experimental results of the falling stick catching. The operator succeeded in catching the object 17 times using the proposed system and 0 times using the 30 fps system. In addition, Figure 19 shows the results of the number of successes and the success rate when the proposed system was used and when the vision system had a latency of 30 ms. It was found that the success rate was 31% higher when the proposed system was used than when the vision was delayed. Figure 20 shows the results of changes in the success rate between the first five times and the second five times when each system was used. In the second half, the success rate tended to increase by 10% for the system with a vision latency of 30 ms. This is considered to indicate that the subject became accustomed to the latency of 30 ms each time and was able to adjust the catch timing. However, even in the latter half, it can be seen that there is a large difference in the number of successes compared to the proposed system. Therefore, it was found that the proposed system was more effective in this experiment than when there was a latency of 30 ms.

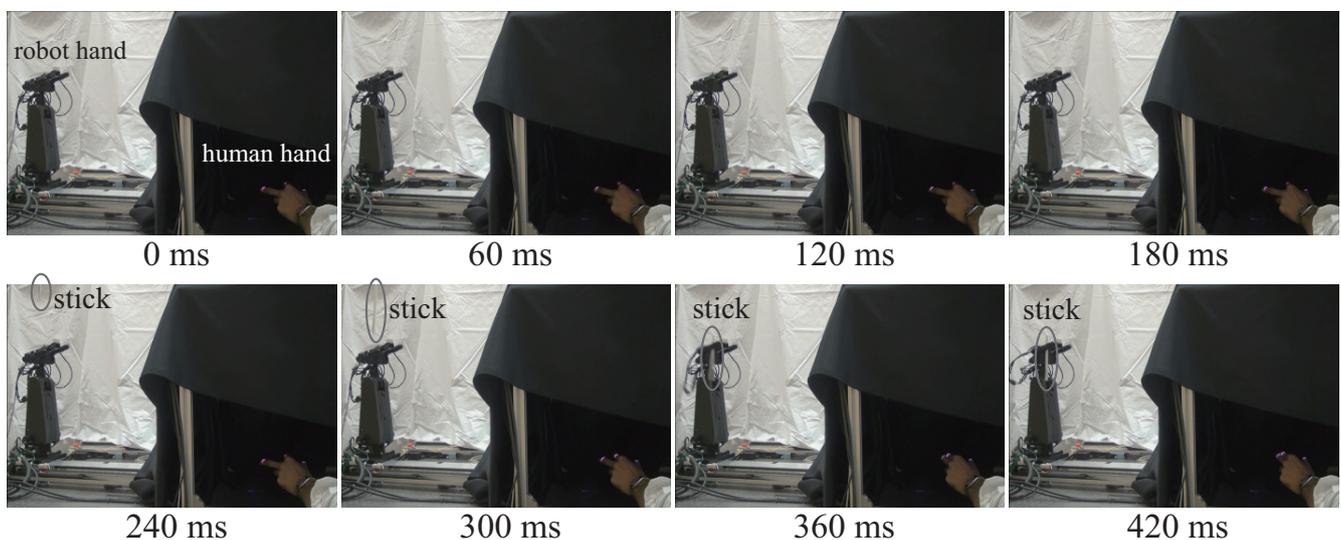


Figure 18. Experimental results of falling stick catching.

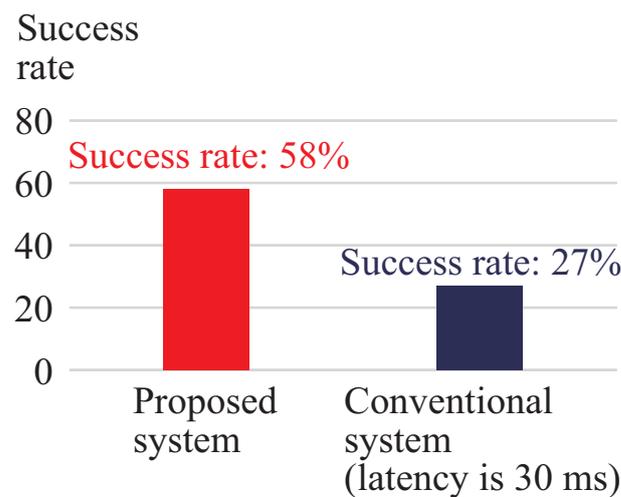


Figure 19. Success rate of falling stick catching.

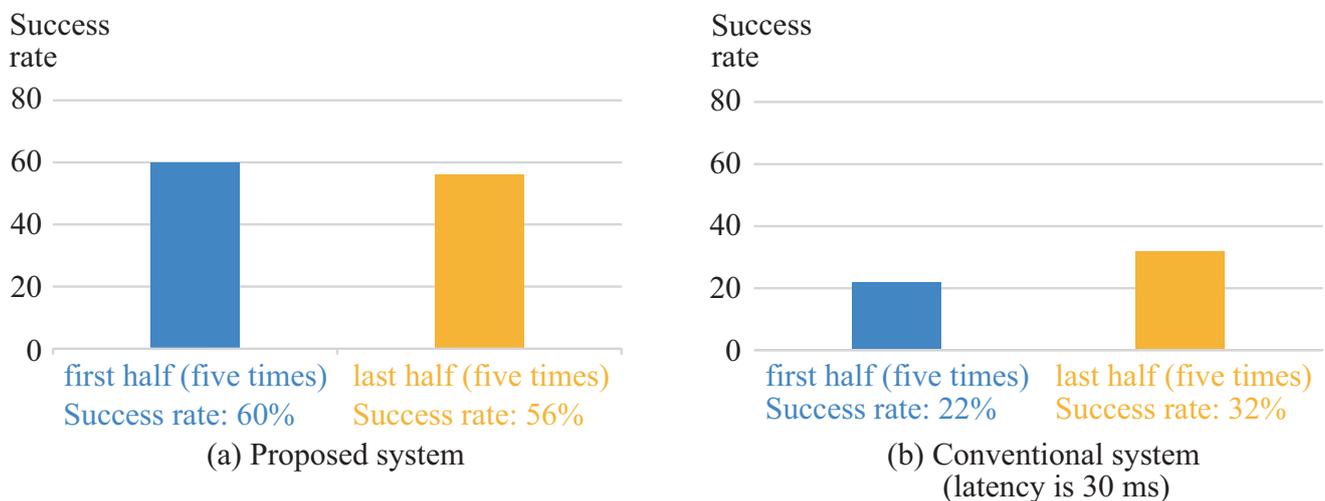


Figure 20. Detailed analysis of success rate of falling stick catching.

6. Conclusions

In this study, we developed an entirely new high-speed, low-latency telemanipulated robot hand system. The developed system is composed of a high-speed vision system including a high-speed projector, a high-speed robot hand and a real-time controller. For the developed system, we propose new methods of 3D sensing of human hand, mapping between the human hand and the robot hand, and the robot hand control with PD control law. After the development of the system, we evaluated the system and demonstrated several experiments using the system. The characteristics and advantages of the developed system are:

- non-contact, high-speed 3D visual sensing of the human hand;
- intuitive motion mapping between human hands and robot hands; and
- low-latency, fast responsiveness to human hand motion.

Such a high-speed, low-latency telemanipulated robot hand system can be considered to be more effective from the viewpoint of usability. In the system evaluation, we confirmed the validity of the proposed method through simulation and discussed the performance (reproduction, latency, and responsiveness) of the developed system through actual experiment with the developed system. As a result, we found that the latency of the developed system was so small that humans cannot recognize the latency. Moreover, we conducted experiments of opening/closing motion as one basic telemanipulated action,

ball catching as one demonstration of static object grasping, and falling stick catching as one demonstration of dynamic object grasping. Finally, we confirmed the validity and effectiveness of the developed system and proposed method.

As future works, we will improve the 3D visual sensing system with more accurately. In particular, we will implement small motion detection method [30] and machine learning based method [31] to the present 3D sensing system. Moreover, we will propose a method of recognizing the intention of the human hand motion, and we will control the robot hand based on the result of the intention. As a result, we will achieve an advanced telemanipulation, which means that the robot hand movements complete before human hand movements complete. By using such a technique, we expect that more distant telemanipulation can be achieved smoothly without human subjects feeling stress due to the high latency. Furthermore, we will extend the present system to a telemanipulated robot hand-arm system to reproduce the human hand-arm motion. In addition, we will integrate softness sensing [32] by introducing a high-speed tactile sensor [33] into the telemanipulated system, and then we will feed back the tactile sensing information to the human hand using a non-contact haptic device [34]. By developing such a system, we will achieve non-contact visual sensing and haptic stimulation in the telemanipulated system at high speed and low latency.

Author Contributions: Conceptualization, Y.Y.; methodology, Y.K., Y.Y. and Y.W.; software, Y.K.; validation, Y.K. and Y.Y.; formal analysis, Y.K. and Y.W.; investigation, Y.K., Y.Y. and Y.W.; writing—original draft preparation, Y.Y. and Y.K.; writing—review and editing, Y.Y., Y.K., Y.W. and M.I.; supervision, Y.W. and M.I.; project administration, Y.Y., Y.W. and M.I.; and funding acquisition, Y.Y. and M.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the JST, PRESTO Grant Number JPMJPR17J9, Japan.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sheridan, T.B. *Telexistence, Automation and Human Supervisory Control*; The MIT Press: Cambridge, MA, USA, 1992.
2. Cui, J.; Tosunoglu, S.; Roberts, R.; Moore, C.; Repperger, D.W. A review of teleoperation system control. In Proceedings of the Florida Conference on Recent Advances in Robotics, Boca Raton, FL, USA, 8–9 May 2003; pp. 1–12.
3. Tadakuma, R.; Asahara, Y.; Kajimoto, H.; Kawakami, N.; Tachi, S. Development of anthropomorphic multi-DOF master-slave arm for mutual telexistence. *IEEE Trans. Vis. Comput. Graph.* **2005**, *11*, 626–636. [[CrossRef](#)] [[PubMed](#)]
4. Angelika, P.; Eienkel, S.; Buss, M. Multi-fingered telemanipulation—Mapping of a human hand to a three finger gripper. In Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, Munich, Germany, 1–3 August 2008; pp. 465–470.
5. Tezuka, T.; Goto, A.; Kashiwa, K.; Yoshikawa, H.; Kawano, R. A study on space interface for teleoperation system. In Proceedings of the IEEE International Workshop Robot and Human Communication, Nagoya, Japan, 18–20 July 1994; pp. 62–67.
6. Arai, S.; Miyake, K.; Miyoshi, T.; Terashima, K. Tele-control between operator's hand and multi-fingered humanoid robot hand with delayed time. In Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, Munich, Germany, 1–3 August 2008; pp. 483–487.
7. Bachmann, E.R.; McGhee, R.B.; Yun, X.; Zyda, M.J. Inertial and magnetic posture tracking for inserting humans into networked virtual environments. In Proceedings of the ACM Symposium on Virtual Reality Software and Technology, Banff, AB, Canada, 15–17 November 2001; pp. 9–16.
8. Verplaetse, C. Inertial proprioceptive devices: Self-motion-sensing toys and tools. *IBM Syst. J.* **1996**, *35*, 639–650. [[CrossRef](#)]
9. Fukuda, O.; Tsuji, T.; Kaneko, M.; Otsuka, A. A human-assisting manipulator teleoperated by EMG signals and arm motions. *IEEE Trans. Robot. Autom.* **2003**, *19*, 210–222. [[CrossRef](#)]
10. Kofman, J.; Wu, X.; Luu, T.J. Teleoperation of a robot manipulator using a vision-based human-robot interface. *IEEE Trans. Ind. Electron.* **2005**, *52*, 1206–1219. [[CrossRef](#)]
11. Chen, N.; Chew, C.-M.; Tee, K.P.; Han, B.S. Human-aided robotic grasping. In Proceedings of the 21st IEEE International Symposium on Robot and Human Interactive Communication, Paris, France, 9–13 September 2012; pp. 75–80.
12. Hu, C.; Meng, M.Q.; Liu, P.X.; Wang, X. Visual gesture recognition for human-machine interface of robot teleoperation. In Proceedings of the 2003 IEEE/RSJ International Conference Intelligent Robots and Systems, Las Vegas, NV, USA, 27–31 October 2003; pp. 1560–1565.
13. Cela, A.; Yebes, J.J.; Arroyo, R.; Bergasa, L.M.; Barea, R.; López, E. Complete Low-Cost Implementation of a Teleoperated Control System for a Humanoid Robot. *Sensors* **2013**, *13*, 1385–1401. [[CrossRef](#)] [[PubMed](#)]

14. Yang, K.; Bergasa, L.M.; Romera, E.; Huang, X.; Wang, K. Predicting Polarization Beyond Semantics for Wearable Robotics. In Proceedings of the 2018 IEEE-RAS 18th International Conference on Humanoid Robots, Beijing, China, 6–9 November 2018; pp. 96–103.
15. Hoshino, K.; Tamaki, E.; Tanimoto, T. Copycat hand—Robot hand generating imitative behavior. In Proceedings of the 33rd Annual Conference of the IEEE Industrial Electronics Society, Taipei, Taiwan, 5–8 November 2007; pp. 2876–2881.
16. MacKenzie, I.S.; Ware, C. Lag as a determinant of human performance in interactive systems. In Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems, Amsterdam, The Netherlands, 24–29 April 1993; pp. 488–493.
17. Katsuki, Y.; Yamakawa, Y.; Watanabe, Y.; Ishikawa, M. Super-Low-Latency Telemanipulation Using High-Speed Vision and High-Speed Multifingered Robot Hand. In Proceedings of the 2015 ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts, Portland, OR, USA, 2–5 March 2015; pp. 45–46.
18. Katsuki, Y.; Yamakawa, Y.; Watanabe, Y.; Ishikawa, M. Development of Fast-Response Master-Slave System Using High-speed Non-contact 3D Sensing and High-speed Robot Hand. In Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–2 October 2015; pp. 1236–1241.
19. Namiki, A.; Imai, Y.; Ishikawa, M.; Kaneko, M. Development of a High-speed Multifingered Hand System and Its Application to Catching. In Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453), Las Vegas, NV, USA, 27–31 October 2003; pp. 2666–2671.
20. Griffin, W.B.; Findley, R.P.; Turner, M.L.; Cutkosky, M.R. Caribration and mapping of a human hand for dexterous telemanipulation. In Proceedings of the ASME IMECE 2000 Conference Haptic Interfaces for Virtual Environments and Teleoperator Systems Symposium, Orlando, FL, USA, 5–10 November 2000.
21. Wang, H.; Low, K.H.; Wang, M.Y.; Gong, F. A mapping method for telemanipulation of the non-anthropomorphic robotic hands with initial experimental validation. In Proceedings of the 2005 IEEE International Conference on Robotics and Animation, Barcelona, Spain, 18–22 April 2005; pp. 4229–4234.
22. Okumura, K.; Oku, H.; Ishikawa, M. High-Speed Gaze Controller for Millisecond-order Pan/tilt Camera. In Proceedings of the 2011 IEEE International Conference on Robotics and Automation, Shanghai, China, 9–13 May 2011; pp. 6186–6191.
23. Janken (Rock-Paper-Scissors) Robot with 100% Winning Rate: 3rd Version. Available online: <https://www.youtube.com/watch?v=Qb5UIPeFCIM> (accessed on 5 January 2021).
24. Telemanipulation System Using High-Speed Vision and High-Speed Multifingered Robot Hand. Available online: <http://www.youtube.com/watch?v=L6Teh3zMSyE> (accessed on 5 January 2021).
25. Huang, S.; Shinya, K.; Bergström, N.; Yamakawa, Y.; Yamazaki, T.; Ishikawa, M. Dynamic compensation robot with a new high-speed vision system for flexible manufacturing. *Int. J. Adv. Manuf. Technol.* **2018**, *95*, 4523–4533. [[CrossRef](#)]
26. Grill-Spector, K.; Kanwisher, N. Visual recognition as soon as you know it is there, you know what it is. *Psychol. Sci.* **2005**, *16*, 152–160. [[CrossRef](#)] [[PubMed](#)]
27. Katsuki, Y.; Yamakawa, Y.; Ishikawa, M. High-speed Human/Robot Hand Interaction System. In Proceedings of the 2015 ACM/IEEE Int. Conference on Human-Robot Interaction, Portland, OR, USA, 2–5 March 2015; pp. 117–118.
28. Ishikawa Group Laboratory. Available online: <http://www.k2.t.u-tokyo.ac.jp/fusion/Janken/> (accessed on 5 January 2021).
29. Laming, D.R.J. *Information Theory of Choice-Reaction Times*; Academic Press: Cambridge, MA, USA, 1968.
30. Asano, T.; Yamakawa, Y. Dynamic Measurement of a Human Hand Motion with Noise Elimination and Slight Motion Detection Algorithms Using a 1 kHz Vision. In Proceedings of the 2019 Asia Pacific Measurement Forum on Mechanical Quantities, Niigata, Japan, 17–21 November 2019.
31. Yoshida, K.; Yamakawa, Y. 2D Position Estimation of Finger Joints with High Spatio-Temporal Resolution Using a High-speed Vision. In Proceedings of the 21st SICE System Integration Division Annual Conference, Fukuoka, Japan, 16–18 December 2020; pp. 1439–1440. (In Japanese)
32. Katsuki, Y.; Yamakawa, Y.; Ishikawa, M.; Shimojo, M. High-speed Sensing of Softness during Grasping Process by Robot Hand Equipped with Tactile Sensor. In Proceedings of the IEEE SENSORS, Busan, Korea, 1–4 November 2015; pp. 1693–1696.
33. Gunji, D.; Namiki, A.; Shimojo, M.; Ishikawa, M. Grasping Force Control Against Tangential Force by Robot Hand with CoP Tactile Sensor. In Proceedings of the 7th SICE System Integration Division Annual Conference, Hokkaido, Japan, 14–17 December 2006; pp. 742–743. (In Japanese)
34. Hoshi, T.; Takahashi, M.; Iwamoto, T.; Shinoda, H. Noncontact Tactile Display Based on Radiation Pressure of Airborne Ultrasound. *IEEE Trans. Haptics* **2010**, *3*, 155–165. [[CrossRef](#)] [[PubMed](#)]