Master's Thesis in Graduate School of
Library, Information and Media Studies

A Study on Embodied Expressions in Remote Teleconference

March 2021
201921627
Kohei Ogawa

A Study on Embodied Expressions in Remote Teleconference
身体性表現を用いた遠隔コミュニケーション会議に関する研究

Student No.: 201921627
氏名: 小川　航平
Name: Kohei Ogawa

I propose a remote communication system which introduced depth sensing technology. I also use it to conduct remote teaching on 3D built objects, and investigated the changes in people's communication. Depth sensors can measure a shape of an object by measuring the distance to objects. However, the use cases of those depth sensors have been utilized only by specialists in special fields. Sensors that recognize human motion, shape, and posture have been shipped in large numbers and have been used in many studies in HCI (Human-Computer Interaction) industry, however they have not been widely used as sensors for general human use. In such a situation, mobile devices equipped with depth sensors have been released and applications and features using the technology have been proposed, and the shipment volume and depth sensing market are about to increase significantly. Considering this background, it is expected that new ways of communicating will be created according to people's communication methods and use cases.

In this study, I focus on the field of telecommunication, starting from the consideration of the object to be photographed, and then I take up two specific use cases to investigate the possibility of new applications of depth sensors for personal users. For that reason, I designed and implemented a telecommunication system using the depth sensor in a videoconferencing system. The system allows the user to change the viewpoint of the image of the caller in the video conference. I also adopted teaching as a task in which the results of communication are clearly visible in order to investigate the changes in people's communication.

The results also showed the interestingness of the teleteaching experience and the effectiveness of the proposed telecommunication system. Similarly, the results showed that remote teaching is difficult even for the task of 3D construction when the method of construction, the method of assembly and the appearance of the assembled object are different.In terms of systems and communication, 70% of participants answered that a viewpoint change function was necessary, and 90% answered that a viewpoint change function would make it easier to communicate intentions and instructions, etc.

In the future, it is necessary to investigate the use case of remote teaching, where real-time and depth data are more required to limit the relative instructions. The discussions and limitations of the system described in this paper will lead to further research into distance communication for personal use cases.

Principal Academic Advisor: Yoichi OCHIAI
Secondary Academic Advisor: Sayan SARCAR

A Study on Embodied Expressions in Remote Teleconference

Kohei Ogawa

Graduate School of Library, Information and Media Studies

University of Tsukuba

March 2021

Contents

List of Figures

Chapter 1

Introduction

1.1 Background

Depth sensors can measure a shape of an object by measuring the distance to objects [1, 2, 3, 4]. These sensors are very useful in areas such as automated driving and the robotics industry, for example, to detect surrounding objects in automated driving, to estimate the position of the vehicle on a map [5], and to detect defects in the shape of objects in production [6, 7, 8]. However, the use of these depth sensors has only been exploited by specialists in specialised fields. In the 2010s, Microsoft launched Kinect[1], a sensor for home game consoles with a built-in depth sensor [9, 10, 11, 12]. In addition to the depth sensor, Kinect has a high-resolution visible light camera and four microphones that can distinguish the direction and type of sound source with high resolution. By recognizing, analyzing, and interpreting the information detected by these multiple sensors, it becomes possible to learn more about the user and the situation in the space. In other words, it is possible to recognize the space in the depth direction, so that the detailed movements of the game player's hands, wrists, arms, legs and hips, as well as the shape of the skeleton and its movements, can be recognized in 3D information.

Such sensors for recognizing human movement, shape and posture have been shipped in addition to Kinect, and have been used in many studies in the HCI (Human Computer Interaction) industry, but they have not been widely used for home gaming [13, 14, 15, 16, 17]. In this context, Apple's launch of tablet devices(iPad Pro[2]) with depth sensors in 2018 and mobile devices (iPhone[3]) with two types of depth sensing technology in 2020, as well as the proposed applications and features using the technology, will significantly increase the shipments of depth sensors and the depth sensing market. Prior to this, general public people had never before had depth sensors for personal use, and this event represents a change in the technological ecosystem that is about to take place. A similar case in the past was the introduction of touch displays and cameras in personal mobile phones (Figure 1.1[18]). For this reason, internet access with simple touch gestures and the increasing use of cameras have created a world where people can easily upload photos to social networking sites and communicate with each other. Considering this background, it is expected that new ways of communicating will be created according to people's communication methods

[1] https://www.microsoft.com/ja-jp/p/azure-kinect-dk/8pp5vxmd9nhq
[2] https://www.apple.com/jp/ipad-pro/
[3] https://www.apple.com/jp/iphone/

and use cases. Therefore, I believe that investigating new use cases when individuals possess depth sensors will lead to the expansion of the depth sensing market and the proposal of new forms of communication.

1.2 Research Motivation

The purpose of this research is to explore new use cases of depth sensing technology for individuals and to investigate the changes in communication associated with them. For this purpose, one of the following situations using depth sensing will be discussed and described in this paper.

In recent years, with COVID-19 raging, face-to-face communication has become a risk of infection. In this context, many people have become accustomed to using video conferencing tools such as ZOOM[4] and Teams[5] to communicate with others. However when you have to communicate the situation in real time in three dimensions, smooth communication is not easy without various perspectives and depth information.

In order to solve such a problem, it would be good to have a tool that can control the video with a free viewpoint by adding depth data to the video of the video conference. In addition, this research will be investigated whether there is a change in the way people communicate remotely when the viewpoint can be changed in three dimensions, while they have only communicated in two-dimensional video.

1.3 Proposal for This Research

In this study, the use cases of depth sensing technology for individuals were explored and the associated changes in communication were investigated. For this purpose, this study was taken as one example of the implementation of depth sensing in telecommunication.

Starting from the consideration of the objects to be measured by the depth sensor, three preliminary experiments and one main experiment were carried out to investigate the changes in communication during the distance. For those experiments, I developed two systems in this study. The first one is a system that uses real-time pose estimation from multi-view camera images to switch the image of itself displayed on the videoconferencing tool. The second is a system that allows the user to control the other person's image in the video conference tool from any viewpoint. Finally, in the main experiment, I used the second implemented tool to create a 3D construction. Specifically, one person was assigned to teach and the other to be taught, and they remotely assembled the blocks. As a result, it was found that it was possible to give a lecture to an object which was partially difficult to detect due to the characteristics of the depth sensor. In the survey questionnaire, 70% of the participants felt that the ability to change the point of view was necessary, and 90% said that the ability to change the point of view made it easier to convey their intentions and instructions.

[4]https://zoom.us/jp-jp/meetings.html
[5]https://www.microsoft.com/ja-jp/microsoft-365/microsoft-teams/group-chat-software

1.4 Contributions

The main contributions of this study are the following:

- Development of software that allows you to change the point of view of the other person's image, and proposal of a video conferencing system that implements this software, its limitations and discussion.
- Discovering and exploring use cases where users can control perspectives in distance communication.
- Investigation of the changes in the discussion of objects and their communication with the characteristics of depth sensors on mobile devices.

Figure 1.1: Camera sales chart showing the explosive growth in smartphone sales among camera types (smartphones are shown by yellow bars).

Chapter 2

Related Work

My study combines insights from depth sensing, switching perspectives and telecommunications. In this chapter, I provide an overview of the previous research works and its relation to my proposed method.

2.1 Depth Sensing for Individual People Applications

In recent years, the demand for depth sensors has been increasing. In the field of HCI, there is a lot of research on the use of Kinect as an input interface for systems. One example is inFORM [19], a dynamic shape display that can physically render 3D content. In this system, a Kinect is placed above the user's head to track the user's hands and surface objects to enable interaction with the system. Song et al. [20] also used Kinect for hand tracking and interaction with images on Mixed Reality. Work is also expected to be done using depth sensors in smartphones to recognise 3D gestures [21].

Depth sensing is also used for face recognition. if only 2D camera images are used for recognition, the system can be fooled by using 2D images of the correct person. Therefore, it has been used in many researches on face recognition using 3D cameras. For more work and discussion on face recognition, I refer to the survey paper by Yassin et al. [22] The integration of face recognition into mobile computers has led to the proposal of several applications. One of Apple's applications is Animoji, which can be used to recognize a user's facial expression in real time. Because it can track the user's facial expressions in real time, it can then replace the expressions of characters such as dogs or robots with its own.

Therefore, it is also being used in Augmented Reality applications, where LiDAR sensors can be used to create 3D models of objects and 3D measurements on real world objects [23].

Finally, papers have been carried out to investigate the accuracy and limitations of digitalised 3D modelling of the own body, such as the feet [24] and face [25], as targets.

2.2 Switching Perspectives

One way to share visual information with people at different points is to transmit images through a camera. If it is possible to switch the point of view of the image, it is possible to obtain a large amount of information about a distant location using only visual information. There are three main ways of switching the image:

1. Method to switch between the images of several cameras
2. Method to switch continuously between images in free view
3. Method to track and switch the image according to the viewpoint

The first method is often used as a television shooting technique. In this way, the video is switched in order to make the content of the image more impressive or to create a change of scene, so that the image does not become monotonous. The job of video switching is considered to be a difficult one, as it requires a higher level of skill and a wealth of experience, which can only be acquired after years of experience as a cameraman. For this reason, there are several studies on the motivation for automatic video switching [26, 27]. For example, Daemen et al. [28] conducted a study in which a robot tracked the movements of actors and switched between images based on cinematographic rules. A study has also been carried out to automatically crop and zoom the actor's face for objects with a large shooting range, such as a stage performance. Kumar et al.[29] proposed a method to generate a video showing a focus+context view based on the context of the scene and the calculated close-up camera movement based on the actor's position data. In the field of video summarization, several methods have been proposed, such as video selection based on object- and frame-level features [30], or video summarization based on converting audio into text and extracting the corresponding frames [31]. In the field of video summarization, several methods have been proposed.

The second method is used to selectively view images from any position or angle, and is used for watching sports [32] and live music videos. There are two methods: one is to reproduce the subject space by modeling the light rays obtained from the camera image as a mathematical formula [33], and the other is to synthesize the image by estimating the amount of displacement of the pixels of the virtual viewpoint by using a depth sensor with infrared rays [34]. However, while these methods enable high-quality image synthesis, they require a large number of cameras, depth sensors and other equipment [35, 36, 37].

In this study, I use a method that converts the depth image into a 3D point cloud and switches the viewpoint of the mapped color image. The only equipment used be the camera and depth sensor on the smartphone.

2.3 Telecommunications

Remote communication is a field of research that has received a great deal of attention during the onslaught of COIVD-19. As people's lives have moved from face-to-face communication to remote communication, many things have been removed. For example, visual information about the other party is greatly affected by the specifications, placement and number of cameras installed on the other party. In addition, the delay associated with the transmission of voice over a network and the input and output devices used can greatly affect the way the sound is heard. It is known that the physical environment for video conferencing is not optimized for the individual, and that the tight schedule due to the lack of travel time increases fatigue. On the other hand, elements of non-verbal communication, such as gestures and facial expressions, have become more prominent because of the difficulty in conveying information to the other party in the environment. With such a thriving change in the way people communicate, there has been a lot of research focusing on this aspect [38, 39, 40].

In this section, I present research on video conferences and remote collaborations that is relevant to this study.

2.3.1 Video Conference

In the field of humanities research, Brid et al. [41] examined how the audio characteristics of video-mediated communication differed from face-to-face interaction, while Marjolijn et al. [42] examined a comparison of the attractiveness of face-to-face communication after chat-based communication and video conferencing communication respectively. Marjolijn et al. [43] also investigated the richness of vocabulary when describing previously unseen pictures in video and audio conferencing conditions.

In the field of display research, OmniEyeball [44, 45] integrates a spherical display system with an omnidirectional camera to capture and display 360° panoramic live streaming video. They then present a unique application for 360° symmetric video communication using two OmniEyeball terminals, and present a potential solution to the problem of narrow field of view in video communication. Kushida et al. [46] developed a telepresence system that uses the screen deformation synchronized with the movement of an object on the projected image, with the aim of extending the videoconferencing system by adding a sense of depth. The results of experiments using the system suggest that the perception of depth enhances the presence of the remote person and the object in the video.

There is also work on the use of Kinect sensors in videoconferencing systems. DeVincenzi et al. [47] proposed a videoconferencing system that blurs out non-speakers and objects by estimating who is currently speaking based on sound and camera data from the Kinect. They proposed a videoconferencing system that blurs people and objects other than the speaker by estimating the current speaker based on sound data from Kinect and camera data.

2.3.2 Remote Collaboration

Information sharing is essential in order to eliminate the limitations of geographical location and to achieve the same goals. In a previous study, IllumiShare [48] proposed a system of surfaces where physical and digital information is shared. It did the job of extending communication through diagramming and sketching, where drawing on physical tables and digital devices is shared. In addition, Holoportation [49] proposed a telecommunication system that enables high-quality real-time 3D reconstruction of the entire space, including people, furniture and objects, which can be transmitted to remote users in real time. This allows the user wearing a virtual or augmented reality display to see, hear and interact with the remote participants in 3D, as if they were in the same physical space. Such a system, however, requires a lot of hardware and processing computers. In addition, recreating face-to-face communication does not always improve the quality of communication.

Therefore, the ability of the robot is employed as a mediator to support communication. The field of them thought that physical body is a powerful conduit for affective communication, and thus mediating properties of the physical body through an agent has been proposed to address this deficiency. Therefore, Saadatian et al. [50] proposed a pair of robots named "Mini-Surrogate" as a personalized telecommunication medium which recreates and emulates the physical presence of a specific person. They applied the concepts of enclothed and embodied cognition to convey the illusion of a particular person's presence. They showed that personalization could enhance communication affectivity in terms of providing awareness and reminding of the specific person.

2.4 Position of This Study

Depth sensing technology has often been used by special people for special applications. However, with the adoption of depth sensors in smartphones, depth sensors are becoming democratised. In this context, depth sensors can play a major role in 3D reconstruction of people and objects in telecommunication. Although tele-communication with depth-enhanced images has been studied, it requires some special hardware and is highly restricted in where it can be used. There have also been some humanities studies on telecommunication, however, none of them have investigated the changes in people's communication using teleconferencing with 3D objects, depth-enhanced camera images and changeable viewpoints. In this study, I implemented a tele-teaching system using smartphones, which are owned by many people. In addition, I am examining the changes in the words used, communication methods, and building time when comparing the different viewpoints. I also investigate the use and experience of two different types of sensors (TrueDepth sensor and LiDAR) in iPhone and iPad during remote teaching. These investigations and discussions will be useful for the introduction of depth sensors in smartphones, for proposing new forms of communication, and for discussing the possibility of remote teaching systems with added depth data.

Chapter 3

Implementation of Real-time Multi-View Video Switching Systems

In this chapter, an implementation of a system that can switch and record video with real-time skeletal estimation for video taken from multiple viewpoints is described. This system can be implemented in video conferencing, and can specifically do the following:

- Multiple camera images can be switched in real time at specific triggers (at regular intervals), and the video files can be saved.
- Each video file arranged in multiple viewpoints and numerical data of estimated skeletal coordinates can be recorded.

3.1 Pose Estimation Using TPU

In this section, the skeletal estimation technique used in this study will be described. Human Pose Estimation technology is a computer vision technology that detects a person in an image or video and can estimate the movement of each joint in the body. This means that the computer can determine where the elbows, shoulders, or feet will appear in the image (Figure 3.1 [51]). Recently, this technique can be used not only for skeletal estimation of 2D coordinates, however also for 3D coordinates. However, in this study, the 2D coordinates of the joint points of a person were estimated for each camera image, considering the accuracy and real-time nature of the estimation. PoseNet [52] used in this study is lighter than other skeletal estimation techniques because it does not recognize the person in the image, but only estimates where the major body joints are located.

PoseNet is a machine learning library that has been improved to work in real time in a web browser. However, currently, it is not possible to keep multiple cameras running by accessing them from a web browser. Therefore, it is necessary to develop an application that can run on the desktop with high real-time performance.

In this study, I use the Coral USB Accelerator[1], which can connect a TPU (Tensor Processing Unit) to a computer (Figure 3.2 [53]).

[1]https://coral.ai/products/accelerator/

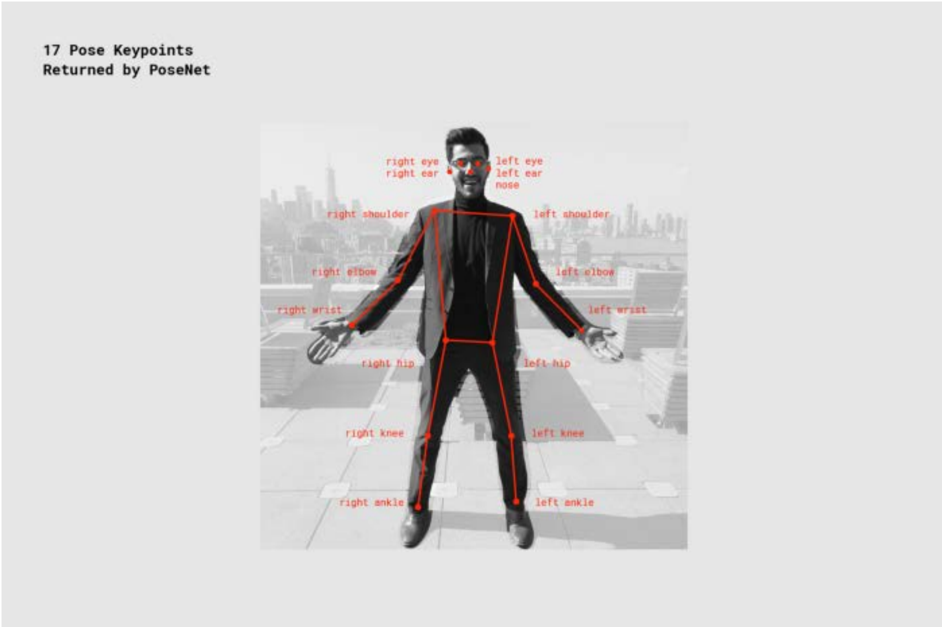

Figure 3.1: KeyPoint estimated by PoseNet.

TPU [54] is a processor that specializes in computing the tensors done in machine learning. By connecting the USB Accelerator that contains it to a compatible Linux computer, machine learning inference can be performed on the device without the need for a CPU or GPU due to its high processing performance.

In addition, the system uses PoseNet models converted to TensorFlow Lite to be able to run on edge TPU [55]. The flow from the TensorFlow model to running on Coral HardWare is shown in Figure 3.3.

In this study, pose estimation is performed in real time using a set of PoseNet models that have been quantized and optimized to run on Google's edge TPU.

3.2 System Overview

In this section, the system configuration and the processing flow of this system will be explained. The system consists of any number of web cameras, a PC, and the USB Accelerator described in the previous section. An example of the camera placement is shown in Figure 3.4. The camera should be placed so that its coverage area at least covers the upper part of the user's body.

The software used is OBS (Open Broadcaster Software)[2], which is used to switch the videos from each camera, and Google Meet[3], which is used as a video conferencing tool

[2]https://obsproject.com/ja
[3]https://meet.google.com/

Figure 3.2: External view of the Coral USB Accelerator.

(any video conferencing tool that can input a virtual camera is possible). The programming language and version used for the development of this system is python 3.7.7[4].

Next, I describe the system processing flow. I used OpenCV[5], an image processing library, to perform the processing related to the input, data processing, and output of this system, and obs-websocket [56, 57] to perform the OBS processing remotely.

First, the system extracts camera devices recognized by the computer. Then, the streaming video is input based on the recognized camera ID using the functions of OpenCV library. The captured videos is processed for skeletal extraction as every frame image, and the lines connecting the skeletal coordinates and joints are plotted on the image. Then, according to the event trigger, the video on OBS is switched by API processing of obs-web-socket. In addition, each camera image, each skeletal coordinate data, and the video during the video conference are written to the local disk.

Each skeletal coordinate data is written out for each frame in the format shown in Figure 3.6. The details are shown below.

Pose： PoseNet returns a pose object containing a list of keypoints and an instance-level confidence score for each detected person.

Keypoints： A part of the person's pose to be estimated, such as nose, right ear, left knee, right foot, etc. It contains both the position and keypoints confidence score. Currently, PoseNet detects 17 keypoints as shown in Figure 3.1 in the previous section.

[4]https://www.python.org/downloads/release/python-377/
[5]https://opencv.org/

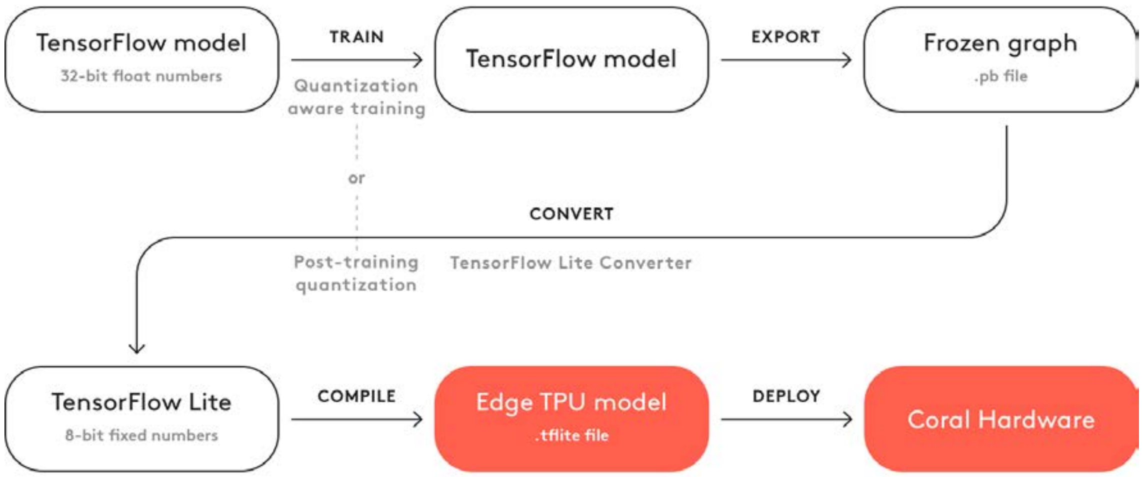

Figure 3.3: The basic workflow to create a model for edge TPU.

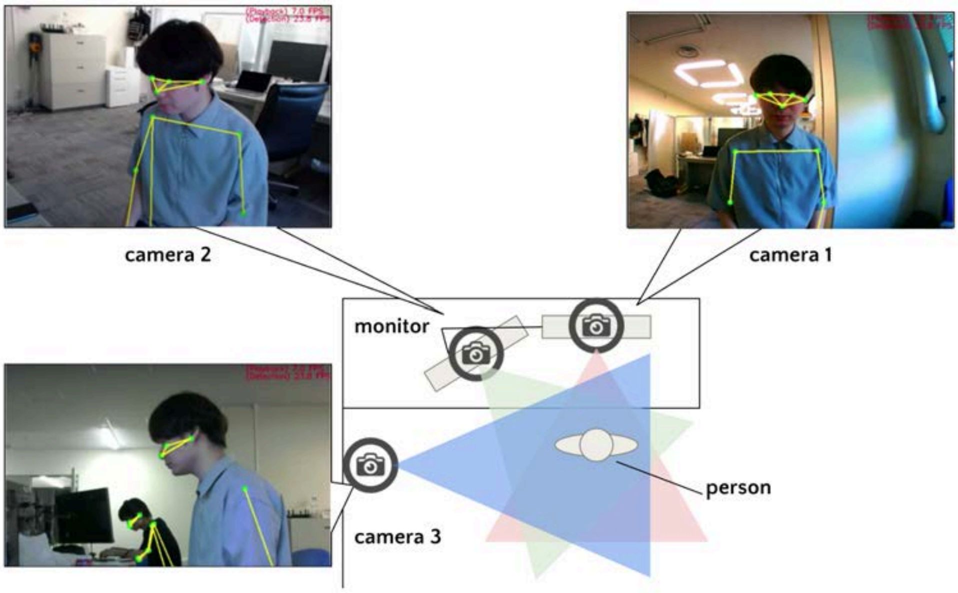

Figure 3.4: Example of camera placement.

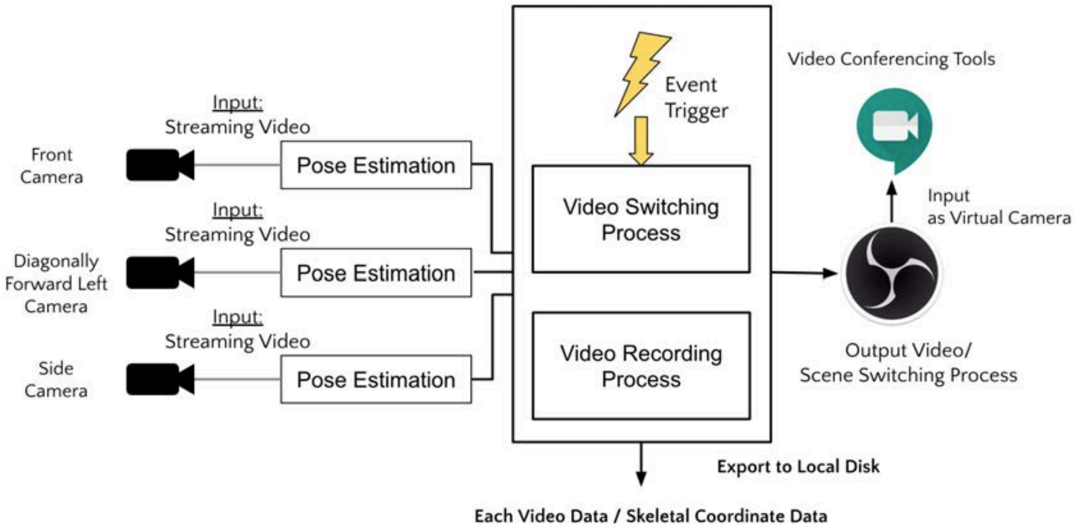

Figure 3.5: System processing flow of real-time multi-view video switching systems.

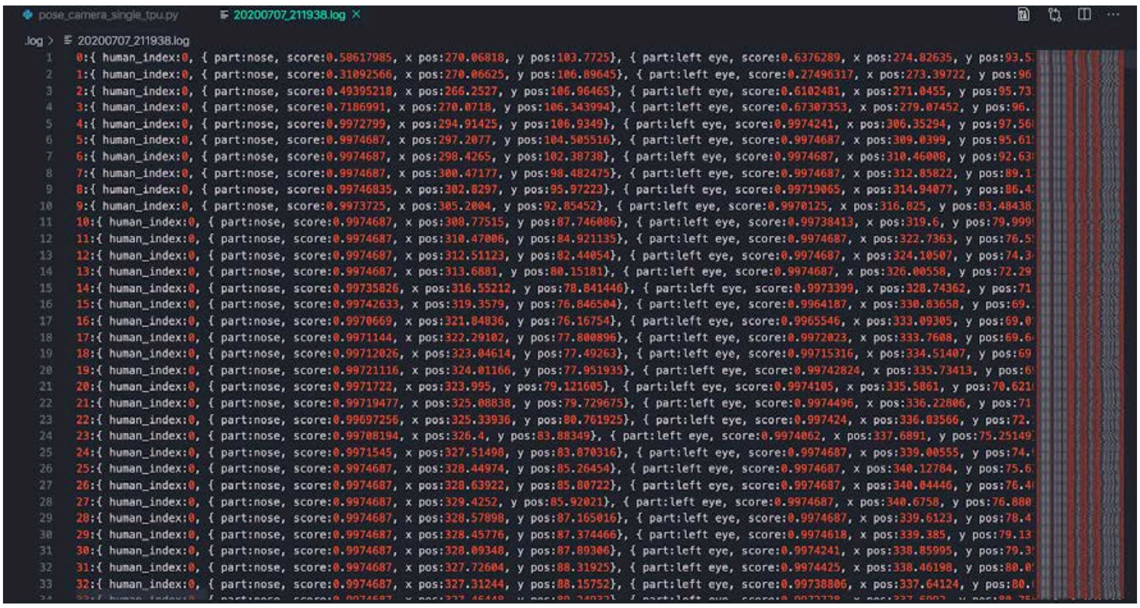

Figure 3.6: Pose estimation results for each person obtained by PoseNet.

Chapter 4

Implementation of Free Viewpoint Video Switching Systems

In this chapter, the implementation of an application that can be switched between different viewpoints using a camera and sensor attached to a mobile terminal, and the design of telecommunication using this application are described.

4.1 Depth Sensors for Apple Device

In this section, depth sensors and sensing methods used in this study, their specific use cases and applications are described. The mobile devices used are iPad Pro [1] and iPhone 12 Pro[2] sold by Apple Inc.

The devices are equipped with depth sensors of two different methods, TrueDepth [58, 59] for the front camera and LiDAR for the rear camera.

4.1.1 True Depth

TrueDepth is made up of multiple cameras and sensors (Figure 4.1). The 3D sensing method is based on structured light method [60], which is a similar technology to Microsoft's Kinect. In structured light method, a pattern consisting of a large number of stripes or arbitrary hinges is projected from a dot projector at a time, and the depth estimation of a large number of samples can be obtained simultaneously. By adopting this method for mobile devices and using TrueDepth, users can take advantage of features such as Face ID, which unlocks the device through facial recognition, and Animoji, which allows users to convert their facial movements and expressions into animations. In short, high performance tracking capability is now available in mobile devices. Using Face ID mechanism as an example, I will explain TrueDepth [61, 62, 63, 64].

1. First, a proximity sensor recognizes the distance between the device and the face.
2. Then, the recognition is sent to TrueDepth system.
3. TrueDepth system then uses the dot projector to project more than 30,000 invisible infrared beams onto the face in order to measure the microscopic shape of the face (Figure 4.2).

[1]https://www.apple.com/jp/ipad-pro/specs/
[2]https://www.apple.com/jp/iphone-12/

4. The infrared light reflected back from the face is captured by the infrared camera and sent to the image processor in the device. The infrared light emitted by the flood illuminator and the image taken by the front camera, which is an RGB camera, are also sent to the image processor so that the face can be identified even in the dark.
5. Then, a bionic chip with a neural engine converts the depth map and infrared images into a mathematical model, which is then checked against the data of the registered face for authentication.

This is the principle on which TrueDepth sensor operates and is utilized.

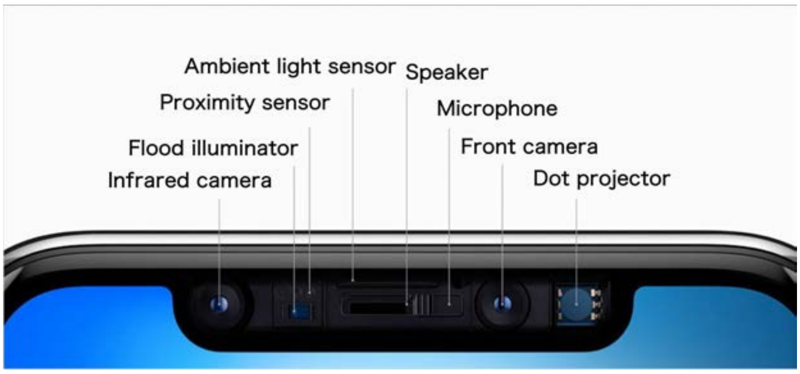

Figure 4.1: Sensors on the front of iPhone 12 Pro[3].

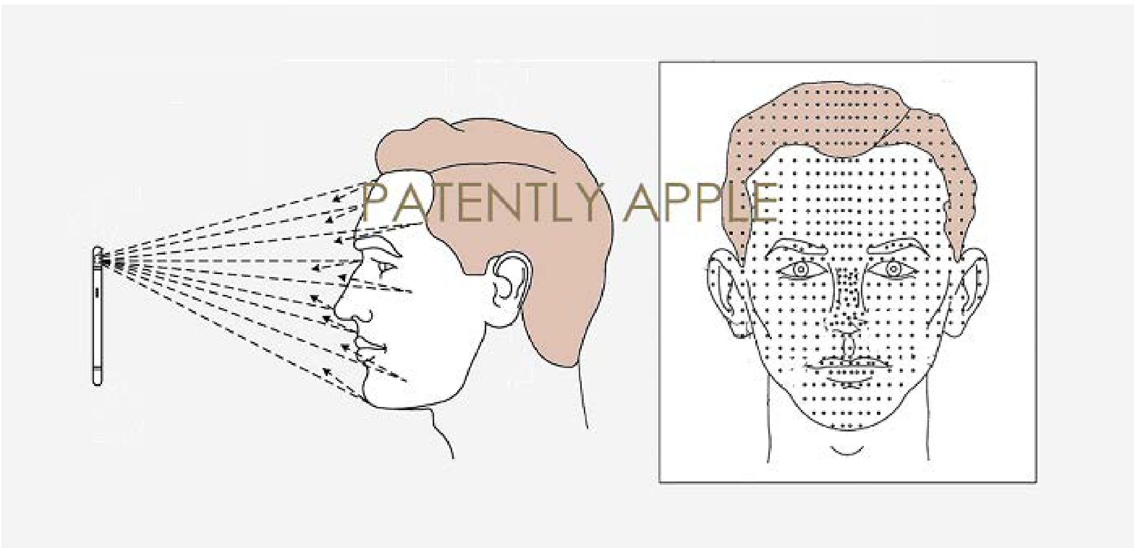

Figure 4.2: Dot projector emitting an infrared beam.

4.1.2 LiDAR

LiDAR (Light Detection and Ranging) is found on the back of apple mobile devices(Figure 4.3 [65]). LiDAR is a sensing technology that uses laser light, and while there are several methods, LiDAR used in this research is based on ToF (Time of Flight) [66, 67].

ToF is a principle that calculates the distance to an object, its shape, etc., by irradiating a laser beam from a light source to the object and measuring the scattered or reflected light [68].

There are two types of ToF.

- dToF (direct ToF) ... A method that calculates the distance to an object by measuring the time it takes for light to be reflected back.
- iToF (indirect ToF)... Calculates the distance from the phase difference of the reflected light.

LiDAR in Apple mobile devices uses the former dToF method, which has the advantage of being resistant to external light and capable of measuring with high resolution even at a distance from the object.

In that device, depth sensing technology is applied to camera shooting technology. Conventional smartphone cameras use machine learning technology to recognize only the subject and blur the rest of the image, but in dark spaces, an RGB camera have difficulty recognizing the subject. Considering this, LiDAR, which uses laser light, has the advantage of being more resistant to darkness and less susceptible to the intensity of ambient light. Therefore, by combining an RGB camera and LiDAR technology, it is possible to recognize the subject in the dark (this is a feature called night portrait). In addition, autofocus in shooting in the dark is fast.

Additionally, depth sensing technology is being applied to the technology of augmented reality (AR). Since LiDAR can measure the positional relationship of surrounding objects, it can recognize the position of walls, floors, furniture, etc., and accurately place objects while keeping the 3D relationship between the real world and the virtual world (AR) consistent.

In chapter 4.1, I described the 3D sensing method of the depth sensor in iPhone and iPad, and specific use case applications for individual users. In the next section, I describe an application that uses and implements the sensors I have just introduced.

4.2 3D TelepreApp

This section is described the design and implementation of 3D TelepreApp, a system for switching the other party's video with a free viewpoint.

4.2.1 Overview

3D TelepreApp is a system that allows the user to remotely switch the other party's video with a free viewpoint by gesture operation of the device. This system has two modes (Figure 4.4 (a)): one is to capture your own video and the other is to control the other's video remotely. The development and operating environment for this system is shown in Table 5.1.

Figure 4.3: Sensors on the rear of iPhone 12 Pro.

Table 4.1: Environment of This System.

Language		Swift 5
Development Environment		Xcode 12.2
Production Environment		iOS 14.0 or more, iPadOS 14.0 or more
Development Devices	Model	MacBook Pro (15-inch, 2019)
	OS	macOS Catalina
	Processor	2.3 GHz 8 コア Intel Core i9
	Memory	32 GB 2400 MHz DDR4
	Graphics	Radeon Pro 560X 4 GB Intel UHD Graphics 630 1536 MB

One or two devices running iOS (or iPad OS) are used in this system. The processing flow of the system is shown in Figure 4.5. The device uses two cameras at the same time, an RGB camera and a depth camera (TrueDepth for front camera use, LiDAR is activated for rear camera use), converts the data as a point cloud, and performs 3D rendering (Figure 4.4 (b)(c)).

The details are shown in Section 4.2.2. In addition, when the application is started on a terminal connected to the same network, it looks for the terminal that is performing the capture and makes the connection. The connected terminal can then be used to remotely change the perspective of the video with the same gesture (Figure 4.4 (d)). The details are shown in Section 4.2.3.

4.2.2 Mode 1: 3D Capturing

In this section, process to visualize depth data in the 3D in Figure 4.5 is explained.

In the viewer, the video is rendered as a point cloud and the viewpoint can be changed by touch gestures as described at the end of this section. The internal processing uses Metal Vertex Shaders to control the geometry, and Metal Fragment Shaders to colorize individual vertices and separate depth and color textures [69]. In this study, the implementation was

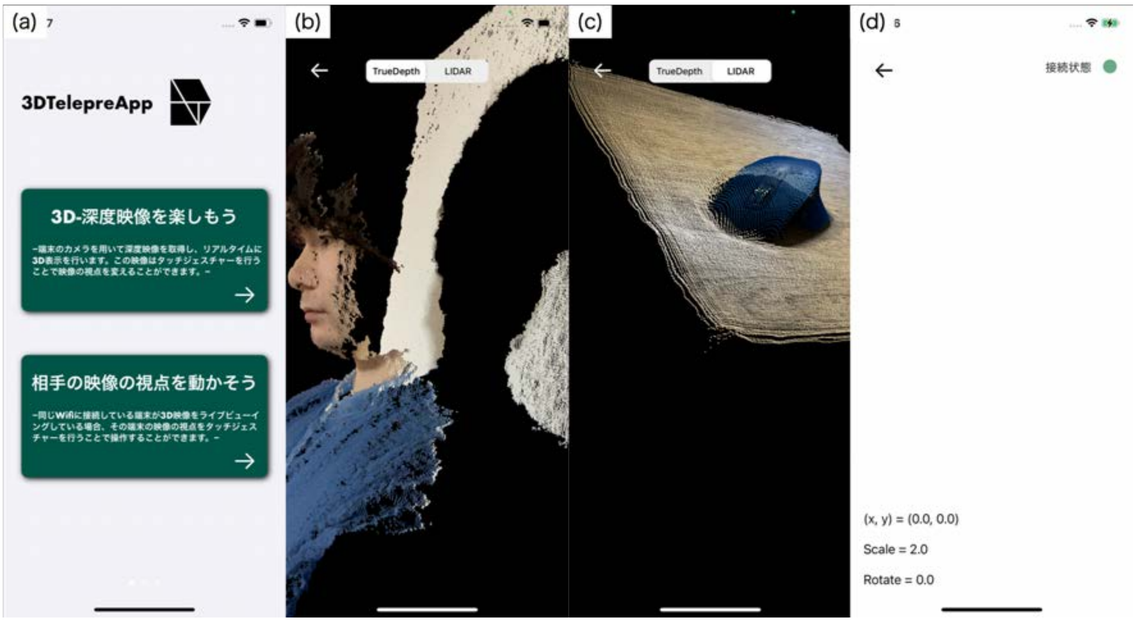

Figure 4.4: 3D TelepreApp UI screen:(a) Top screen, (b) Screen of mode1 (using TrueDepth), (c) Screen of mode1 (using LiDAR), (d) Screen of mode2.

based on the sample code available on the official Apple website [70]. It was implemented in Objective-C, but I rewrote all the source code to work in Swift, which is currently a tributary of Objective-C. In addition, the sample uses AVFoundation[71] framework to acquire a RGB data and a depth data, but in this study, I did not use AVFoundation at all, however used only ARKit[72] framework to stream the video and convert the depth data into 3D. It is not possible to run ARKit and AVFoundation functions simultaneously. Also, switching between using those two frameworks is a problem that takes several seconds in the current environment. In addition, as AR applications using depth sensors are developed in the future, video processing using only ARKit will be indispensable. In this study, I also streamed the LiDAR video and converted the depth data into 3D. Next, the conversion from depth map to 3D point cloud will be explained.

Converting Depth Maps to 3D Point Cloud Coordinates:

In order to change the viewpoint, the depth map obtained from the depth camera needs to be converted to point cloud coordinates (X, Y, Z) in 3D space (Figure 4.6 [73]).

Let the value of the depth map be Z and the coordinates in that map image be (U, V). Since the Z of the point cloud coordinates in 3D space is identical to the value of the depth map, I need to calculate the remaining X and Y.

X and Y can be calculated by using intrinsicMatrix [74] as follows:

$$X = (U - o_x) * Z/f_x \tag{4.1}$$

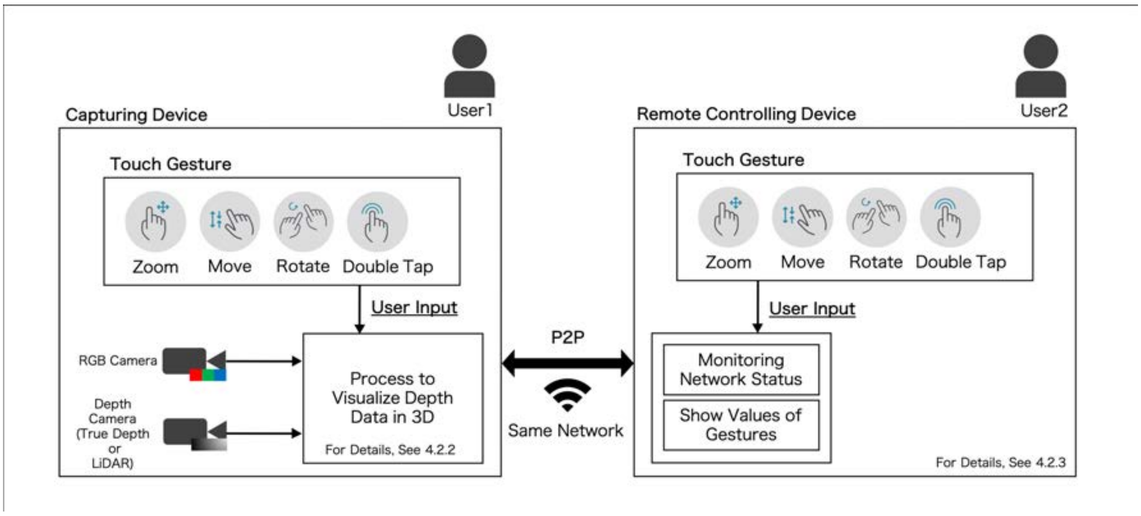

Figure 4.5: System processing flow of 3D TelepreApp.

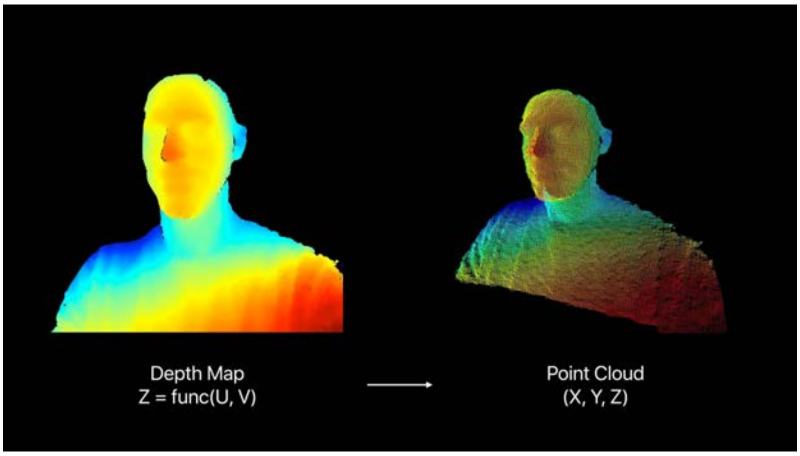

Figure 4.6: Converting depth maps to 3D point cloud coordinates.

$$Y = (V - o_y) * X/f_y \tag{4.2}$$

The intrinsicMatrix [74] is a 3x3 matrix representing the internal parameters of the camera, which are specific to the camera. The f_x,f_y represent the focal length, and o_x,o_y are the offsets of the principal points from the upper left corner of the image frame. All values are expressed in pixels. In other words, with the principal point at the center, the depth value Z (unit: meter) and the focal length of the camera f_x,f_y (unit: pixel) are used to transform X,Y into the coordinate system of real space (unit: meter) for rendering.

Touch Gestures:

The 3D point cloud rendered in all sections can be used to change the viewpoint by making the following touch gestures through the view.

- Pan to move the camera around the center (Figure 4.7 (a)).

- Pinch to zoom (Figure 4.7 (b)(c)).
- Rotate with two fingers to turn the camera angle (Figure 4.7 (d)).
- Double-tap the screen to reset the initial position (Figure 4.7 (e)).

4.2.3 Mode 2: Remote Controlling

In this section, a remote controlling device in Figure 4.5 is described. The user can use the same touch gestures to remotely control the rendered 3D point cloud view with RGB data and depth data (Figure 4.7). In order to do this, the device that acquires the video and the device that controls it remotely must be connected to the same network, and signals are sent through P2P (Peer to Peer) communication. When communication is established, the upper right color in Figure 4.4 (d) is green, and when communication is not possible, it is red. In this system, Multipeer Connectivity Framework [75] is used for the communication part. The flow of the network communication process is shown in Figure 4.8.

4.3 Telecommunication System

In the previous section, I described the design and implementation of a system called "3D TelepreApp", which allows users to switch the other party's video with a free viewpoint.

In this section, a configuration for teleconferencing using the system is proposed. In this study, ZOOM[4] was used as a remote video conferencing tool. In this study, the following screen sharing method was used to display 3D TelepreApp video on zoom. I use a method to from an iOS or iPad device, connect to the USB port via an HDMI capture device, and select "Contents from 2nd Camera" under "Share screen" to display. This ensures that there is as little degradation of resolution and delay in the shared screen as possible.

The configuration of the video distribution and the video displayed at that time are shown in Figure 4.9.

[4]https://zoom.us/jp-jp/meetings.html

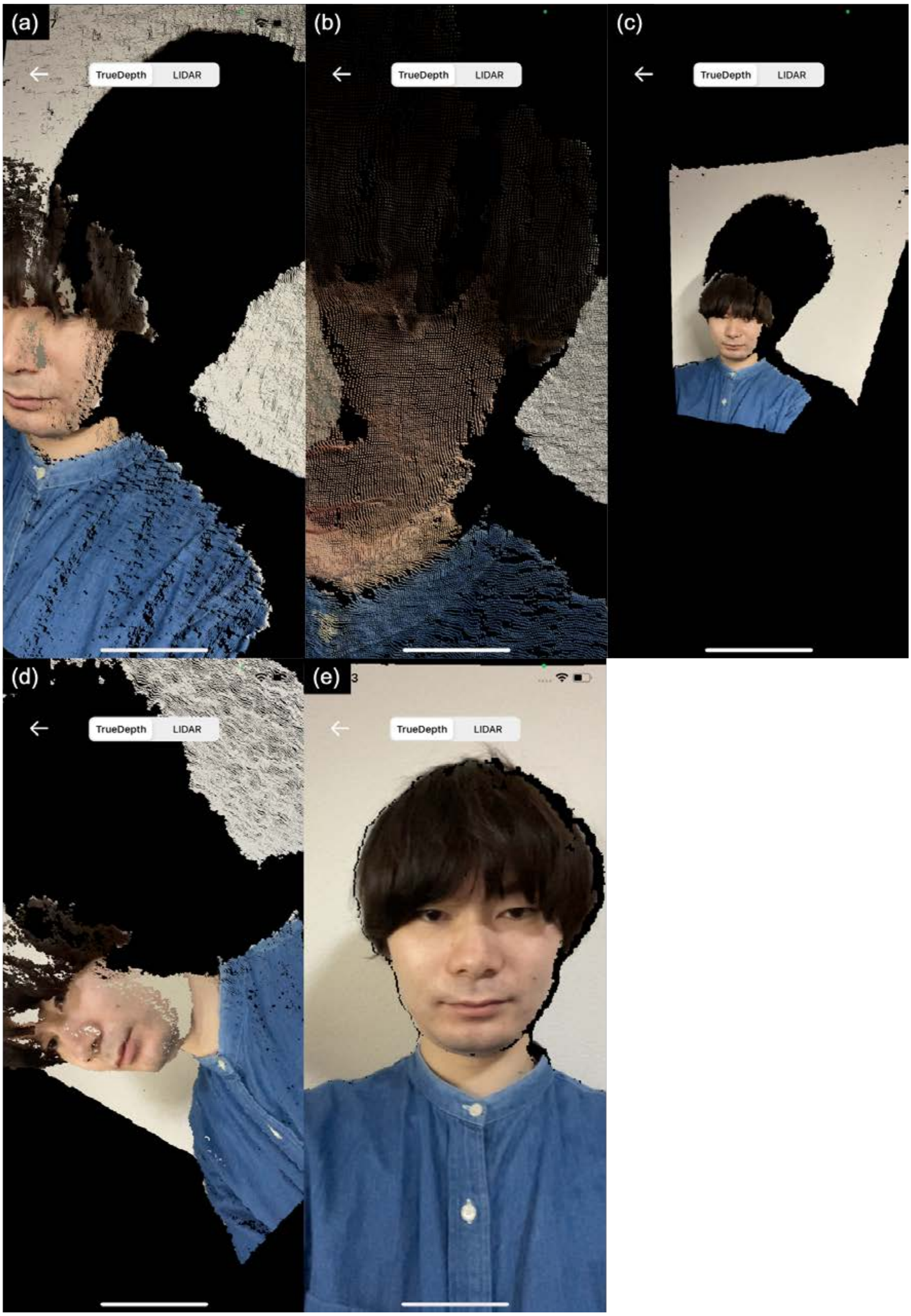

Figure 4.7: Touch gestures:(a) Move, (b) Zoom-In, (c) Zoom-Out, (d) Rotate (e) Reset.

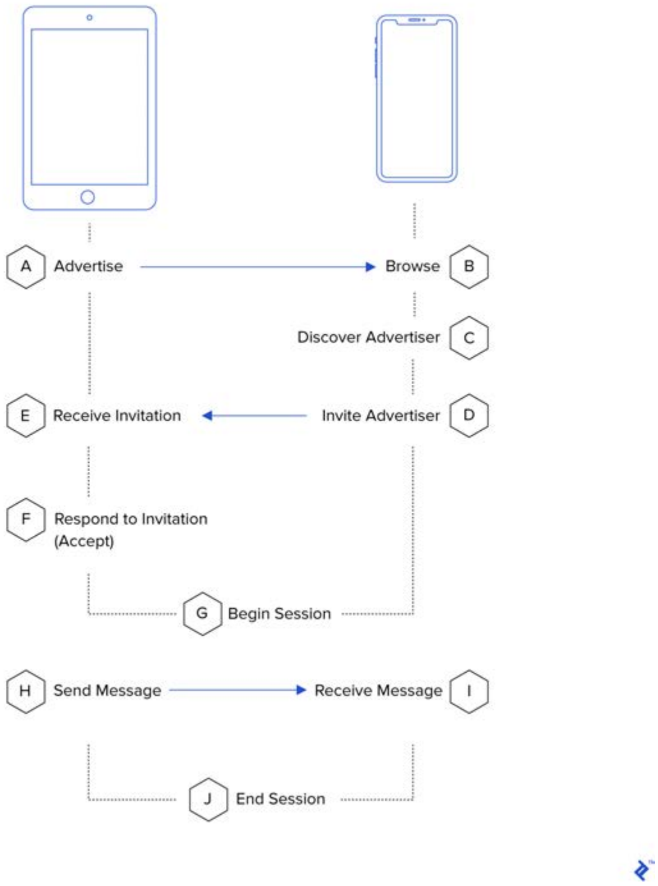

Figure 4.8: Network communication flow of Multipeer Connectivity.

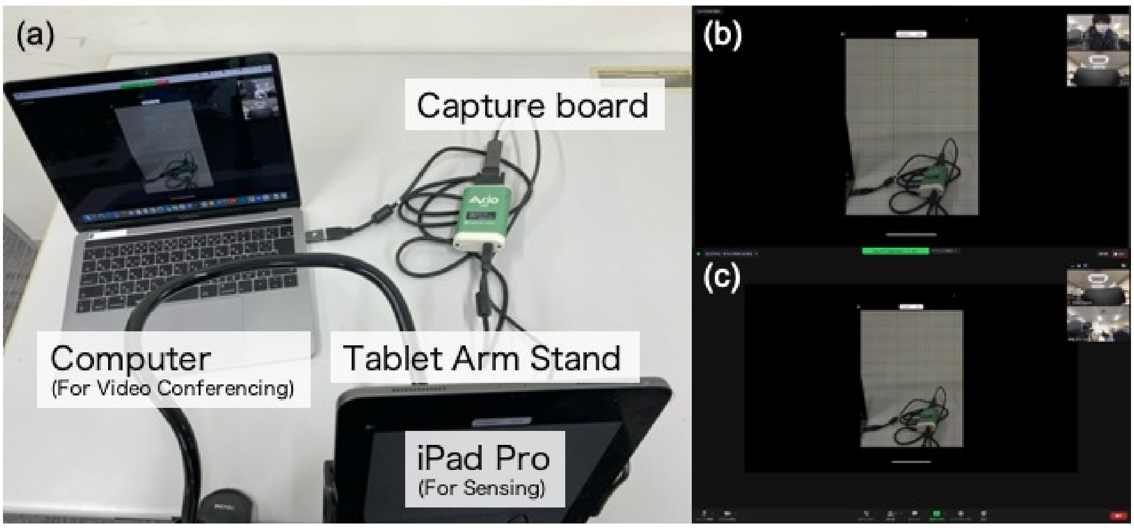

Figure 4.9: (a) Remote communication system configuration, (b) Projected videos on the capture device side, (c) Projected videos on the remote controller device side.

Chapter 5

Pilot Study

The purpose of this study is to investigate "how the usage and communication method will change when depth sensor is introduced into the videoconferencing tool, and what other use cases there are.

The objects to be filmed during video conferencing can be categorized into bodies and objects. When the target is a person, non-verbal communication such as body language and subtle human expressions can be conveyed. On the other hand, when it comes to objects, it can convey the scale, texture, color, and behavior of the object. As described in Section 4.1, the method and characteristics of 3D sensing also differ depending on the depth sensor used. Although two sensing schemes are discussed in this study, it would be a useful contribution to the field of a depth sensing to know whether an object is sensed or not considering the use case of 3D sensing technology in a video conferencing tool.

In this study, the first section of the paper investigated the "representation of physicality in video conferencing" for human subjects.

5.1 Study1: Investigation of 3D Body Movements During Video Conferencing

In this section, the first preliminary experiment is described. In Study 1, a user survey is conducted in order to investigate how the relationship between two parties affects the speaker's body movements in video conferencing.

5.1.1 Participants

Four participants (4 males, all of whom were members of our laboratory) in the 19–24 years age group (M=22.25, SD=1.92) participated in the user survey. Participants were paired in pairs to avoid duplication of age relationship combinations.

5.1.2 Experimental Design

The participants were asked to chat for 15 minutes using the video conferencing tool in the filming environment shown in the next section. This section is intended to provide an overview of the experimental environment and procedures.

Table 5.1: Environment of This System.

Model	Field of View	Resolution	Frame Rate
MacBook Pro (15-inch, 2019)[1]	54 ° [76]	1280 × 720	20fps
C920 n[2]	78 °	1980 × 1080	30fps
CMS-V43BK-3[3]	150 °	1280 × 720	30fps

Table 5.2: Specifications of the web camera used in Study 1.

Position	What to Expect
Front	Changes in facial expression
Overlook	Changes in overall body movement, changes in facial orientation
Side	Changes in the back and forth swaying of the body, changes in gestural movements, and changes in reactions to the screen

Setup

In Study1, the experiment was conducted using the system shown in Section 3. The information of the web cameras used is given in Table 5.1. The placement and expectation of the cameras are in the three directions shown in Table 5.2, and their appearance is shown in Figure 5.1.

Figure 5.1: Camera layout and experimental environment of Study 1.

Procedure

After 15 minutes of free conversation, the participants answered a questionnaire. During the experiment, the video was recorded by a camera placed in the room. In the questionnaire, after investigating the frequency of communication between the two parties and the relationship between them, the start time of the video with body movement and the reason for such movement were investigated with reference to the video recorded by the OBS (including audio). The video recorded by OBS is one in which multiple camera images are

[1]https://www.apple.com/jp_smb_3500/shop/buy-mac/macbook-pro
[2]https://www.logicool.co.jp/ja-jp/product/hd-pro-webcam-c920n
[3]https://www.sanwa.co.jp/product/syohin.asp?code=CMS-V43BK-3

Table 5.3: Questionnaire items for Study1.

Number	Question
1	What is your relationship with your partner?
2	How often do you talk to your partner per week?
3	How tired do you feel after a conversation?
4	Please look back at the video of the conversation and write down the times in the video when there was physical movement and why you moved the way you did.
5	What did you feel because of your relationship with the other person?
6	When did you feel that you should be able to see the other person's physical information?
7	What did you feel from the other person's video?
8	Were there times when you felt that you wanted to show the other person your own image? If yes, when?
9	What do you look at when you talk during a video conference?
10	What is the reason for Question #9?

switched at regular intervals. The questionnaires answered by the experimental participants are shown in Table 5.3.

5.1.3 Result

The relationship between the two parties consisted of one pair of seniors and one pair of juniors who talked about once a week, and one pair of seniors and one pair of juniors who met almost for the first time. In the former combination, the senior and the junior responded that they were not very conscious of how they were perceived because they were the ones they usually talked to, and that they talked to each other in a relaxed manner as in face-to-face communication. On the other hand, juniors to seniors said that they were more tired than usual because they had to put a lot of effort into telling the other person what they were going to say. The junior responded that he felt he was more excited than usual, perhaps because he was conscious of being filmed by the camera, or perhaps because gestures are difficult to convey in a video conference. In the latter combination, the senior to the junior responded that they felt like they could listen carefully to the discussion because the information was fresh. On the other hand, the juniors to seniors said that they could not ask deep questions because they had never met before, but they were looking forward to getting to know each other. In the latter combination, they also said that they did not feel tired.

No one said that they sometimes wanted to show their own images to the other person during the video conference. However, there were times when they wanted the other person to see something, such as a gesture or movement, or when they wanted the other person to listen to what they were saying. The environment of the video conference was also thought to have an effect, with respondents stating that "if you are supposed to be sitting in a chair, you are under different movement constraints than if you are standing" and "it is assumed that your body movements differ depending on the distance from the screen. During the experiment, body movements were estimated posture-wise, but only hand gestures were found to move, not the entire body. Also, assuming that a tool that can change the other person's point of view is introduced, the object to be sensed may be more suitable for the size of the movement of objects or gestures rather than the body.

5.2 Study2: Investigation of Remote Teaching on 3D Construction(Task:Christmas tree decoration)

In this section, the second preliminary experiment will be described. In Study 1, there were not many body movements during the video conference. When considering the use case of changing the viewpoint, it was assumed that the use case of approaching the subject with more movement, such as the movement of a person's hands or an object, would be more effective. Therefore, in Study 2, the subject was changed from a person to a Christmas tree, and the purpose of the experiment was changed to evaluate teaching in a video conference with a 3D object.

5.2.1 Participants

Four participants (1 female and 3 males, all of whom were members of our laboratory) in the 20–25 years age group (M=22.75, SD=1.92) participated in the user survey. In Study2, participants were paired up and assigned the roles of teacher and taught. They attribute the frequency of using video conferencing tools to 2-3 days per week, and face-to-face communication to "the frequency of face-to-face communication per week in Corona Peril (2020.04 or later)" to 4-5 days per week or more.

5.2.2 Experimental Design

Participants were asked to decorate a Christmas tree (tabletop size: 50 cm)[1] using a video-conferencing tool in the filming environment shown in the next section, with each person in charge of teaching and being taught how to decorate. This section is described the experimental environment and the experimental procedure.

Setup

In Study 2, the experiment was conducted using the system shown in Chapter 4. The Christmas tree was placed at a distance of 38 cm from the device for capturing to the center of the tree. That is the maximum distance that can be accommodated by the screen size of the iPad Pro was taken into account. In addition, the iPad Pro screen was set up in such a way that the Front image would show up well on the screen. The prepared decorations are shown in Table 5.2.2. The exterior view of the Christmas tree pre-decorated in this experiment is shown in Figure 5.2. The following items were also pre-decorated in this experiment because the position of the decorations is fixed.

- Christmas Signage
- LED Lights
- Star

[1] https://www.amazon.co.jp/gp/product/B07ZRBGSPR

Table 5.4: List of Christmas tree decorations.

Name	Number of items
Drum	2
Pinecone	2
Cane	2
Christmas Signage	1
Gift Box	2
Candy	2
Bowknot	2
LED Lights	1
Star	1
Bell	2

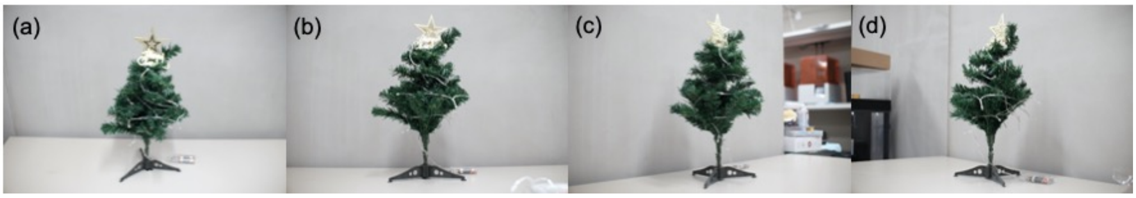

Figure 5.2: Christmas tree in the initial state of the decorating task:(a) Diagonally front, (b) Front, (c) Diagonally left, (d) Diagonally Right.

Procedure

Prior to the execution of the decorating task, the following was explained in advance to the experimental participants (those who taught and those who were taught)

- Type, appearance, and number of items to be displayed.
- Decorating order
- The area to be decorated should be the front half of the tree only.
- Some ornaments are hung on branches, some are placed, and some are decorated in the same way.
- In the case of an experiment with a viewpoint change, both the participant who teaches and the participant who is taught should be able to operate their respective devices.
- The image view on the device is changed to a free viewpoint, and the participant can move the viewpoint, zoom in and out, rotate, and return to the initial position by touch gestures.

The order of the decorations is the order shown in Table 5.2.2 (excluding for those already decorated). The decorations are arranged on the desk in the order in which they are to be displayed. This is so that the task of decorating does not involve spending time looking for the location of the items to be decorated. The person teaching the decorating is provided with an image that shows the order of the decorations. The person who teaches the decorations is given placement instructions based on the correct pattern images from four directions (Diagonally front, Front, Diagonally left and Diagonally right) so that the position of the decorations is known. A list of the correct pattern images used in this experiment is shown in Figure 5.3.

Figure 5.3: List of correct answer pattern images given to the teacher.

Before the experiment, participants were first asked to answer a questionnaire about the attributes of the participants shown in Table 5.2.2. During the task, an experimental assistant measured the time required to complete the tree. The case where the viewpoint is not changed is denoted as Task-A, and the case where the viewpoint is changed is denoted as Task-B. In this experiment, after Task-A was performed, the positions of the teacher and the taught were switched, and Task-B was performed. In this experiment, the verification was also conducted using in-camera (TrueDepth) (Exp.1) and out-camera (LiDAR) (Exp.2). Survey questionnaires were administered every time Task was completed. The questionnaire items for Exp.1-A and Exp.1-B are shown in Table 5.2.2. The questionnaires for Exp.2-A and Exp.2-B are the same. After answering the questions about Exp. 2, the experiment was completed by answering the following questions for the entire experiment.

- What other use cases have you found this system (free viewpoint remote system) to be useful for?

The combination table of participants, correct pattern images, and tasks for Study 2 is shown in Table 5.2.2.

5.2.3 Result

The working time differed from pair to pair, but for Exp. 1, Task-A was faster than Task-B. However, for Exp. 2, Task-A took the same amount of time as Exp. 1-B, while Exp. 2-B

Table 5.5: Questionnaire to investigate attributes in Study2.

Number	Question
1	Age
2	Gender
3	How often do you use the video conferencing application in a week?
4	How often do you communicate face-to-face per week in the midst of a COVID-19 pandemic (2020.04 or later)?
5	Which of the following is the most common use of teleconferencing applications?
6	Which of the following is the second most common use of teleconferencing applications?

Table 5.6: Main questions for Study2.

Number	Role	Question
1	Common	How long did it take you to decorate?
2	Teacher	Was it easy to see the video through zoom when you were teaching?
3	Teacher	Was it easy to teach?
4	Teacher	Please answer this question only if you answered "It was a little easy" or "It was very easy" in the above questions about the experience. Please tell us what you felt was easy.
5	Teacher	Please answer only if you answered "somewhat difficult" or "very difficult" to the above questions about the experience. Please tell us what you found difficult about the experience.
6	Teacher	Which of the following is the second most common use of teleconferencing applications?
7	Common	How much did you enjoy the experience of teaching remotely?
8	Common	Please answer only if you answered "a little fun" or "a lot of fun" to the above questions about the experience. Please tell us what you enjoyed about the experience.
9	Common	Please answer only if you answered "I didn't enjoy it" or "I didn't enjoy it at all" in the above questions about the experience. Please tell us what you felt was not enjoyable and what could be improved.
10	Student	Which did you look at more often, the PC (zoom screen) or the tablet (app screen)?
11	Student	Was the teaching method easy to understand?
12	Student	Was the instruction easy to understand?
13	Student	Please answer this question only if you answered "It was a little easy" or "It was very easy" in the above questions about the experience. Please tell us what you felt was easy.
14	Student	Please answer only if you answered "somewhat difficult" or "very difficult" to the above questions about the experience. Please tell us what you found difficult about the experience.
15	Common	How much did you enjoy the experience of teaching remotely?
16	Common	Please answer only if you answered "a little fun" or "a lot of fun" to the above questions about the experience. Please tell us what you enjoyed about the experience.
17	Common	Please answer only if you answered "I didn't enjoy it" or "I didn't enjoy it at all" in the above questions about the experience. Please tell us what you felt was not enjoyable and what could be improved.

Table 5.7: Combination table of subjects and correct pattern images.

Participant	Exp.1				Exp.2			
	Task-A		Task-B		Task-A		Task-B	
	Teacher	Student	Teacher	Student	Teacher	Student	Teacher	Student
a	Pattern A			Pattern B	Pattern C			Pattern D
b		Pattern A	Pattern B			Pattern C	Pattern D	
c	Pattern B			Pattern C	Pattern D			Pattern A
d		Pattern B	Pattern C			Pattern D	Pattern A	

took more time. This may be due to the fact that they did not know how to give instructions to the Task during the first task. In addition, from the answer that the participants moved their viewpoints to get a better view during the experiment, it is thought that they needed time to move their viewpoints in order to confirm the place to be placed and the place where it was placed. As for the experience of lecturing remotely, some respondents said they enjoyed it a little because "multiple people solved the problem," "both people could move the same thing even if they were far apart," and "the message was conveyed. As for the system, I received the following positive responses.

As for the system, I received the following positive responses

- "It was easy to talk about depth, and it was easy to understand the location of the other person's instructions."
- "The ability to change viewpoints and zoom made it easier to teach."
- "I wanted this system because the pursuit of perfection requires many viewpoint changes."

In terms of the cleanliness of the LiDAR and TrueDepth images, many respondents said that TrueDepth was cleaner because it had fewer black grid lines. However, in terms of operability, many respondents said that TrueDepth was difficult to operate with gestures because it was attached to the front of the device. Therefore, some respondents said that LiDAR was preferable overall.

Other use cases for this system were answered as follows:

- "When the attending physician remotely directs the treatment of the doctors."
- "When detailed work or workmanship is being evaluated."
- "Critique of a sculpture."
- "Assembly work when you have to look at the back."

Other comments included the following

- There were times when the tablet did not display correctly depending on the hand or object, so it was difficult to explain the placement.
- Just thinking that the viewpoint could be changed made me feel a little more secure.

- There were times when I felt a strong delay during the experiment.

The above responses indicated the necessity of changing viewpoints during lectures and the experience of remote lectures. On the other hand, some of the Christmas tree decorations were not displayed and the lecture was troublesome due to the network delay during the experiment.

5.3 Study3: Investigation of Remote Teaching on 3D Construction(Task:Building blocks)

In this section, the third preliminary experiment will be described. In Study 2, I used the system shown in Chapter 4 to give a lecture on decorating a Christmas tree through a video conferencing tool. As a result, although there were some decorations and shapes that could not be detected correctly due to the characteristics of the depth sensor, it was possible to give a lecture on decorating. In addition, the system that can change the viewpoint was able to get positive responses by using the tool with added depth information. However, since there were many complications caused by the inability to sense, the object was changed to building blocks in Study 3 in order to further explore use cases with other objects. In addition, because of the network latency during the experiment in Study 2, the experiment was conducted in a different location with an Ethernet connection. In addition, in Study 2, no restrictions were placed on the teaching method, so the completion speed differed and the individual's teaching ability had a significant impact. Therefore, in this study, instructions for assembling the building blocks are given based on the instructions, and evaluation is conducted. The purpose of the experiment in this section is also to evaluate teaching in video conferencing for 3-D building blocks.

5.3.1 Participants

One participant (1 male, he was members of our laboratory) participated in the user survey. He was twenty-four years old. In Study2, participants were paired up and assigned the roles of teacher and taught. He uses video conferencing applications 4-5 days per week, and face-to-face communication is at least 2-3 days per week for "face-to-face communication per week with Corona Peril (2020.04 or later).

5.3.2 Experimental Design

Participants were asked to assemble building blocks with an experimental assistant using a videoconferencing tool in the filming environment shown in the next section. This section describes the experimental environment and the experimental procedure.

Setup

In Study 3, as in Study 2, the experiment was conducted using the system described in Chapter 4. The PC used in the experiment was the same as in Table 4.1. The starting point for placing the first building block was marked 47 cm back from the capturing device, which is the maximum distance that can be accommodated on the iPad Pro screen according to the molding size of the building blocks to be assembled. It was also placed so that the screen of the iPad Pro would show a good image of the Front.

The building blocks used in this experiment are shown in Figure 5.4.

Figure 5.4: Building blocks in the initial state of the stacking task.

Procedure

As in Study 2, the experiment is conducted with Task-A for the case where the viewpoint is not changed and Task-B for the case where the viewpoint is changed. The verification was conducted for a front camera (TrueDepth) case (Exp.1) and a rear camera (LiDAR) case (Exp.2) as well. Before conducting the experiment, participants were given a questionnaire about the attributes of the participants shown in Table 5.2.2 as in Study 2. During the task, the experimental assistants measured the time required to complete the building blocks. In addition, the experimental assistant will give instructions to the person being taught based on the instructions for assembling the blocks of Appendix described at the end of this paper. There are four patterns in all. In this experiment, as in Study 2, I possess images of the correct patterns from four directions (Diagonally front, Front, Diagonally left and Diagonally right). Figure 5.4 shows a list of the correct pattern images used in this experiment. The building blocks are returned to the state shown in Figure 5.5 every time the assembly task is performed. The survey questionnaire was answered after each task. In addition to the questionnaire for Study 2 (Table 5.2.2), the questionnaire was designed to determine which is better, based on the cleanliness of the image and the reality of TrueDepth or LiDAR.

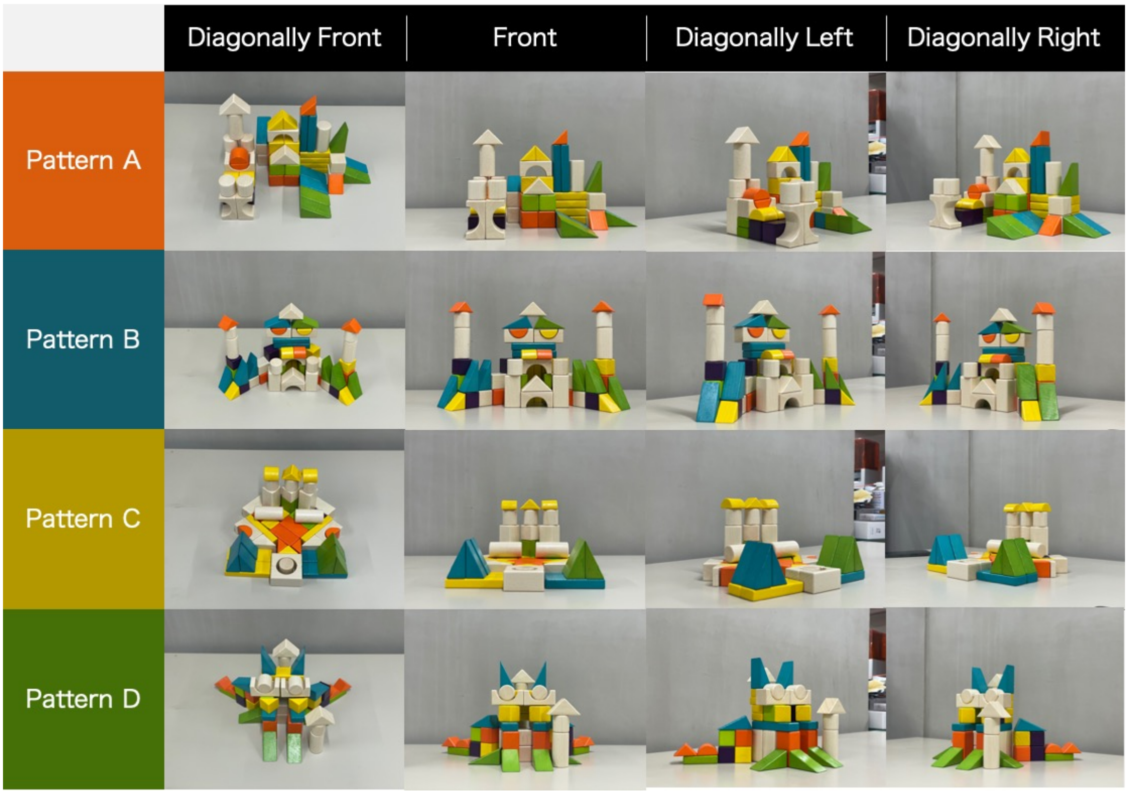

Figure 5.5: List of correct answer pattern images given to the teacher.

5.3.3 Result

For the study, the participants looked at the building blocks throughout the experiment, with or without changing their viewpoint. In addition, almost no viewpoint change was made. Almost no difference in time was observed for all experimental items.

Because the assembly could be done by listening to the voice of the instructor reading out the instructions, and because the camera was placed from the front and the blocks were designed from the back, the situation requiring a change of viewpoint was not created. Also, in this experiment, the respondents said that LiDAR was better. The reason for this was that LiDAR is attached to the out-camera, so the assembly work can be done while watching the image on the tablet. Therefore, they answered that the in-camera system would not motivate the person being taught to change. In addition, with the current camera arrangement, there were many areas where the side of the building blocks could not be seen even if the viewpoint was changed. On the other hand, when the teacher was explaining, changing the viewpoint made it easier to understand the depth direction. Finally, another use case was described as sculpture.

Chapter 6

User Study

The purpose of this research is to investigate whether there is any change in language, communication methods, and remote cognitive abilities over video conferencing when a tool that allows users to change their viewpoints freely in real time is introduced to video conferencing tools. It is also to investigate the use cases and usage of the communication method. in previous chapters, three preliminary experiments were conducted. The first was an investigation of human communication with a focus on physical expression. In the second, a Christmas tree decorating task was conducted by introducing a viewpoint change tool, and the effectiveness of the tool in 3D construction, the possibility of lecturing, and lecturing with depth data were investigated. In the third study, as in the second study, the viewpoint change tool was introduced and the object was changed to a block of wood, and the teacher remotely created 3D objects based on the instructions. These three preliminary experiments confirmed the following:

- The experience of remotely teaching 3D modeling is fun because of the complexity of the content of the objects and the fact that multiple people can solve the problem.
- Considering the operability of the depth sensing on the iPad Pro, it is better to use LiDAR on the out camera.
- Depending on the combination of the object and its camera placement, it was easier to teach by changing the viewpoint.
- The lecture itself is possible even if the object is difficult to detect with the ToF sensor used.
- The viewpoint change is not used unless the object is a construction object or its design that needs to know the depth.

In this experiment, based on the above, the following changes were made compared to Study 3 (Section 5.3).

- The object is still a block of wood, but the camera placement is changed from front to diagonally front.
- Changed the design of the building blocks to require more knowledge of the depth.
- The instructions were eliminated and the participants in the experiment were divided into groups of two.

- Changed to provide practice time to do the same task as in the experiment.

The experiment was then used to evaluate the completeness of the building blocks, the ease with which information was conveyed and communicated, and the change in language used as a result of the introduction of the viewpoint change tool. In this section, the attributes of the subjects, the experimental design and the results will be described.

6.1 Participants

Ten participants (two females, eight males, six of whom were members of our laboratory) participated in the user survey in the age group of 19 to 23 years (M = 22.8 and SD = 1.77).

6.2 Experimental Design

Participants were divided into two groups, one to be taught how to assemble the blocks and the other to be taught how to assemble the blocks, using a videoconferencing tool in the filming environment shown in the next section. In this section, the experimental environment and the experimental procedure will be described.

6.2.1 Setup

In this experiment, the system shown in Chapter 4 was used. The PC used in the experiment was the same one as in Table 4.1. The building blocks used were the same as in Study 3 (Section 5.3), but only the yellow, green, and blue blocks were used (Figure 5.4). In addition, in order to keep the sensing area unchanged in this experiment, the black stand was placed in a fixed position and the area to be assembled was limited. The screen of the iPad Pro was set up so that the diagonally front image would show up well as shown in Figure 6.1(a)(the tilt angle was 45 °). In the horizontal direction, the sensor position of the LiDAR was positioned so that the center of the sensor was aligned with the center of the platform. The environment during the experiment is shown in Figure 6.2.

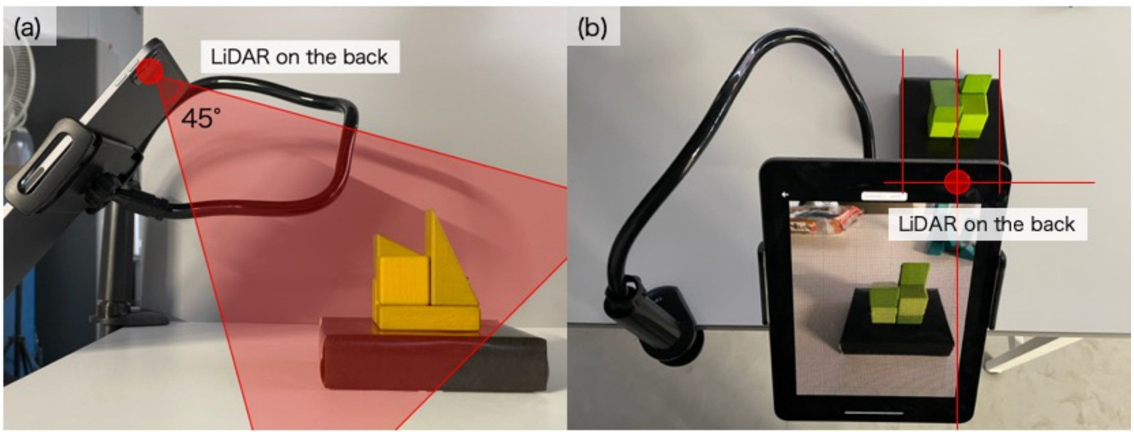

Figure 6.1: Setup for this experiment:(a) Device angle, (b) Location of the device's LiDAR sensor.

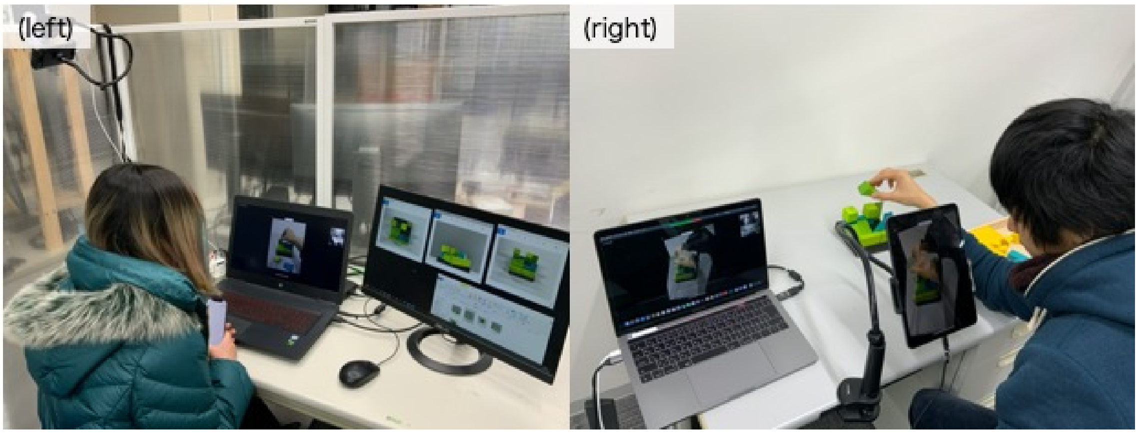

Figure 6.2: Experimental setup and scene:(left) Setting up the teaching side, (right) Setup for the person being taught.

Design of building blocks: Through the video conference images, the correct pattern images need to be created, taking into account the designs of the building blocks that need to change their viewpoints. This was defined as a design where it is not possible to determine which blocks or positions are visible from one direction, or where the blocks appear to be of a different shape. In such a design, the length of the depth is not known without changing the viewpoint. An example is shown in Figure 6.3. The blocks marked in Figure appear to be cubes, but when the viewpoint is changed, the blocks are shown to be oblique in the depth direction as shown in Figure 6.3 (left). The building blocks in this experiment were designed to be funded like this impossible figure.

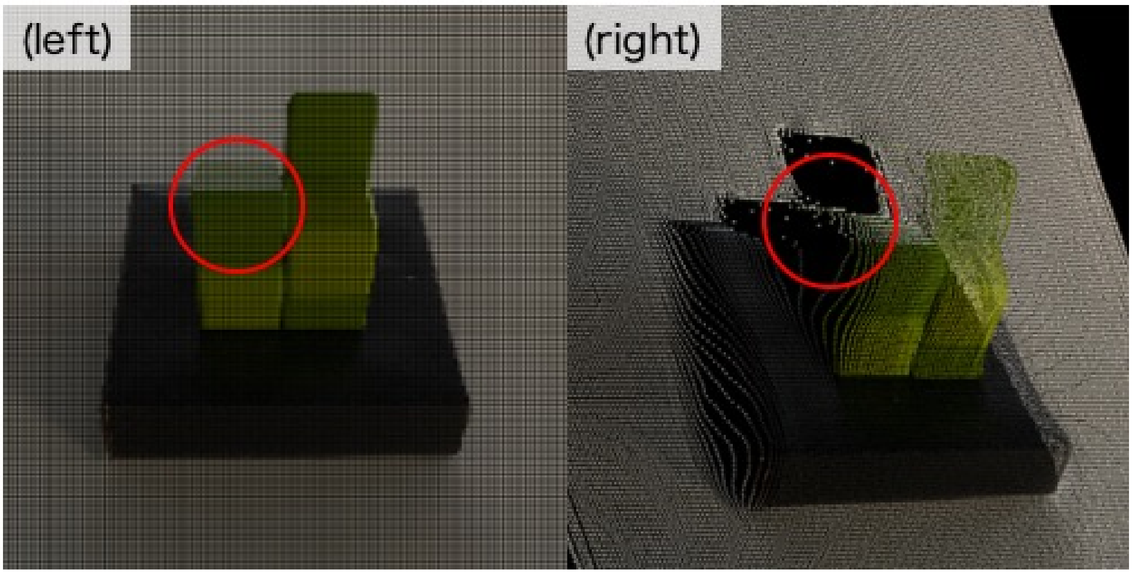

Figure 6.3: An example of a building block design:(left) The inscribed blocks appear to be cubes, (right) Changing the viewpoint shows that it is an oblique side.

6.2.2 Procedure

As in Study2 (Section 5.2) and Study2 (Section 5.4), the experiment is considered Task-A when the viewpoint is not changed, and Task-B when the viewpoint is changed. Only LiDAR on the rear camera was used for depth sensing. Before the accumulation task was performed, the following was explained to the experimental participants (both the teacher and the taught) in advance.

- Each time a task begins, put it back in the box and sort it by color.
- The shape and color of the building blocks to be used (three colors)(Light green and dark green are considered the same color).
- Student should not know the number of building blocks to be used in one task.
- If participants manipulate their viewpoints at the same time, it is difficult to know which one of them did the manipulation, so it is necessary to communicate through conversation.

A box of color-coded building blocks is placed on the desk. In addition, the teacher is instructed to place the completed blocks based on the correct pattern images from five directions (Diagonally Front, Front, Diagonally Left and Diagonally Right, Above).

Before conducting the experiment, participants were first asked to answer a questionnaire about the attributes of the participants shown in Table 5.2.2. During the task, an experimental assistant measured the time required to complete the blocks. The experimental participants were instructed that the teaching participant should make a final check of the building blocks assembled by the taught participant, and that the teaching participant should judge that the building blocks were OK and inform the experimental assistant. In this experiment, the First Task was changed according to the pair of participants in the experiment. The combination table of participants, correct pattern images, and tasks in this experiment is shown in Table 6.2.2. In this experiment, after the completion of the preliminary questionnaire, the participants practiced the sequence of the experiment. The design of the building blocks used in the exercise is shown in Figure 6.4. The correct images used in the exercise were Pattern A, Pattern B, Pattern C, and Pattern D, which were used in the order Task-A, Task-B. The correct pattern images used in the production are shown in Figure 6.5. he survey questionnaire was completed after each Task.

The survey questionnaire was divided into items to be answered for each role and task. Those who taught were asked about the pictures they referred to most often in the correct pattern images, the ease with which the instructions were conveyed, what they tried to convey to others, and the words they used most often. On the other hand, in the questionnaire for Task-A, the person being taught was asked about situations where it was difficult to understand the other person's words, the way the other person communicated, and words that often left an impression. In the questionnaire for Task-B, in addition to the questions in Task-A, the teacher was asked about the timing of the change of perspective and what he or she was thinking about when changing the perspective. After the completion of the Task-A and Task-B experiments, the participants were asked to answer a summary questionnaire. The questionnaire asked the participants what they liked and disliked about teaching and being taught through videoconferencing, how they conveyed their words with and without the viewpoint change, how easy it was to convey their intentions and instructions, and finally, about other use cases.

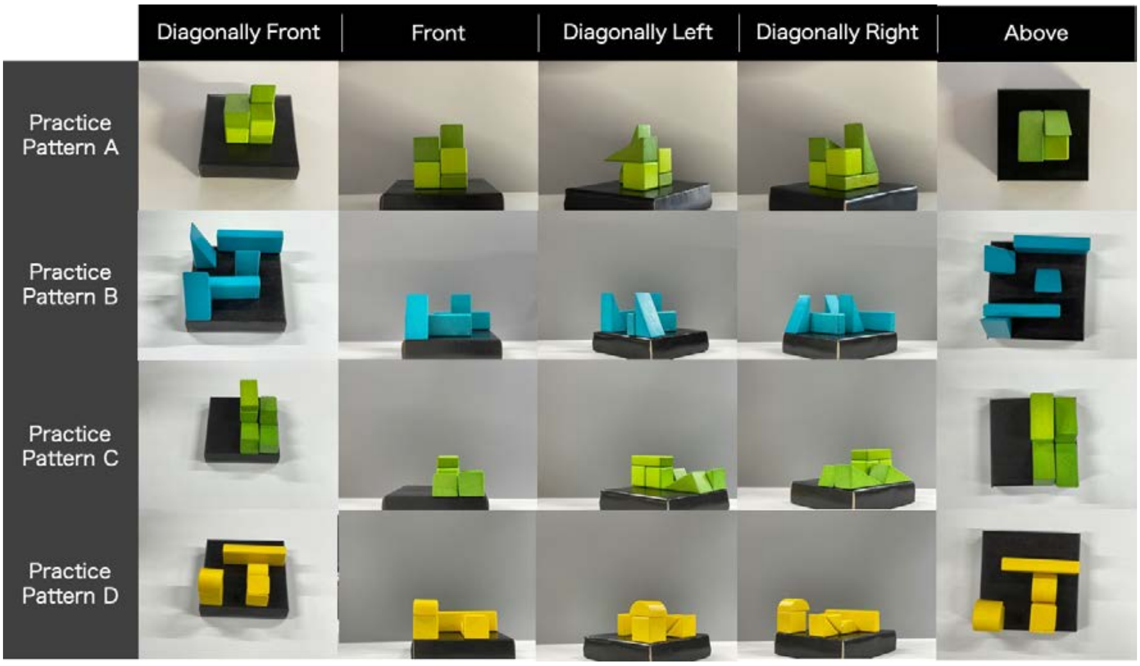

Figure 6.4: List of design patterns of building blocks used in the exercise task.

6.3 Result

The following results were obtained from the communication between the participants during the experiment. Regardless of whether the viewpoint was changed or not, the participants first put the blocks on the floor, and then showed them to the camera and moved them around to get closer to the correct answer. In addition, the teacher named the shapes of the objects as similar to familiar ones to communicate their intentions, and the taught began to use the same names. The change of perspective was often used to get a general picture and to understand the position of the placed parts.

When the correct pattern of the blocks and the actual assembled blocks were checked for consistency, only one out of 20 patterns had the wrong orientation of one part placement. Figure 6.6 shows that almost all the results show that it took longer to assemble the blocks with the viewpoint change than without the viewpoint change. Figure 6.7 shows that people paid more attention to the image directly above the correct pattern in the case without viewpoint change than in the case with viewpoint change. Figure 6.9 shows that "kosoado" word was used more in the case with the viewpoint change. Figure 6.11 shows that 40% of the teachers reported that they changed their perspective when looking at the whole blocks. Figure 6.12 shows that 70% of the respondents said that the ability to change the viewpoint was necessary. Figure 6.13 shows that in the case of a change of perspective, 60% of the respondents said that the language used by the teacher was easy to understand. Figure 6.14 shows that in the case of no change of perspective, 40% of the respondents said that the language used by the teacher was easy to understand. Figure 6.16 shows that 90% of the

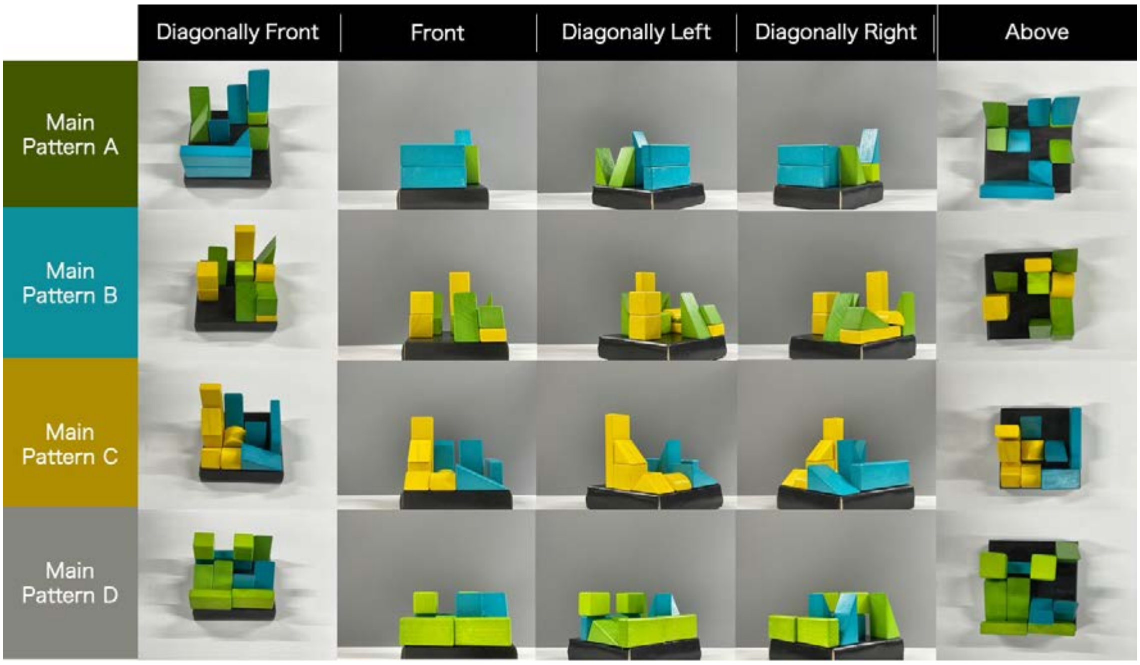

Figure 6.5: List of design patterns of building blocks used in the main Task.

respondents answered that it was easier to communicate their intentions and instructions when they could change their viewpoints.

These results show that the ability to change the viewpoint in remote teaching is effective. In addition, the introduction of the depth sensor enables the design and implementation of a remote communication system, which shows that remote teaching of 3D structures is an effective new application of the depth sensor.

Table 6.1: Combination table of subjects and correct pattern images.

Participant No.	First Task	Task-A		Task-B	
		Teacher	Student	Teacher	Student
1	Task-A	Pattern A	Pattern B	Pattern C	Pattern D
2		Pattern B	Pattern A	Pattern D	Pattern C
3	Task-B	Pattern B	Pattern C	Pattern D	Pattern A
4		Pattern C	Pattern B	Pattern A	Pattern D
5	Task-A	Pattern C	Pattern D	Pattern A	Pattern B
6		Pattern D	Pattern C	Pattern B	Pattern A
7	Task-B	Pattern D	Pattern A	Pattern B	Pattern C
8		Pattern A	Pattern D	Pattern C	Pattern B
9	Task-A	Pattern A	Pattern C	Pattern B	Pattern D
10		Pattern C	Pattern A	Pattern D	Pattern B

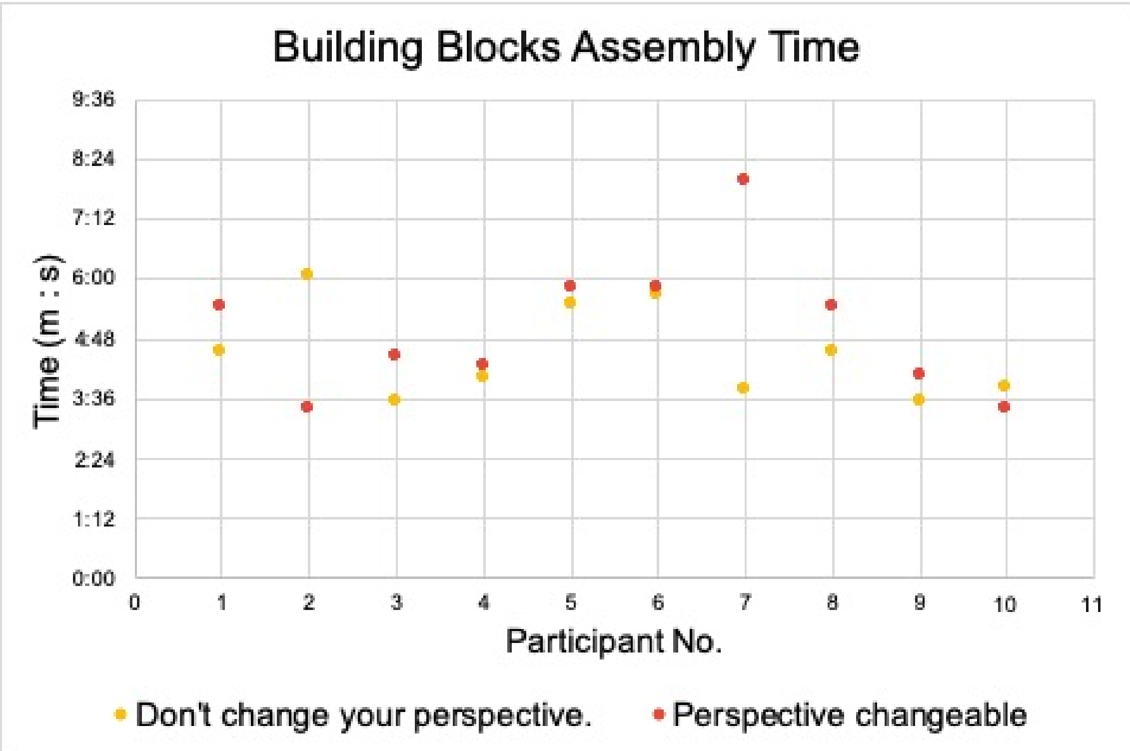

Figure 6.6: Building Blocks Assembly Time.

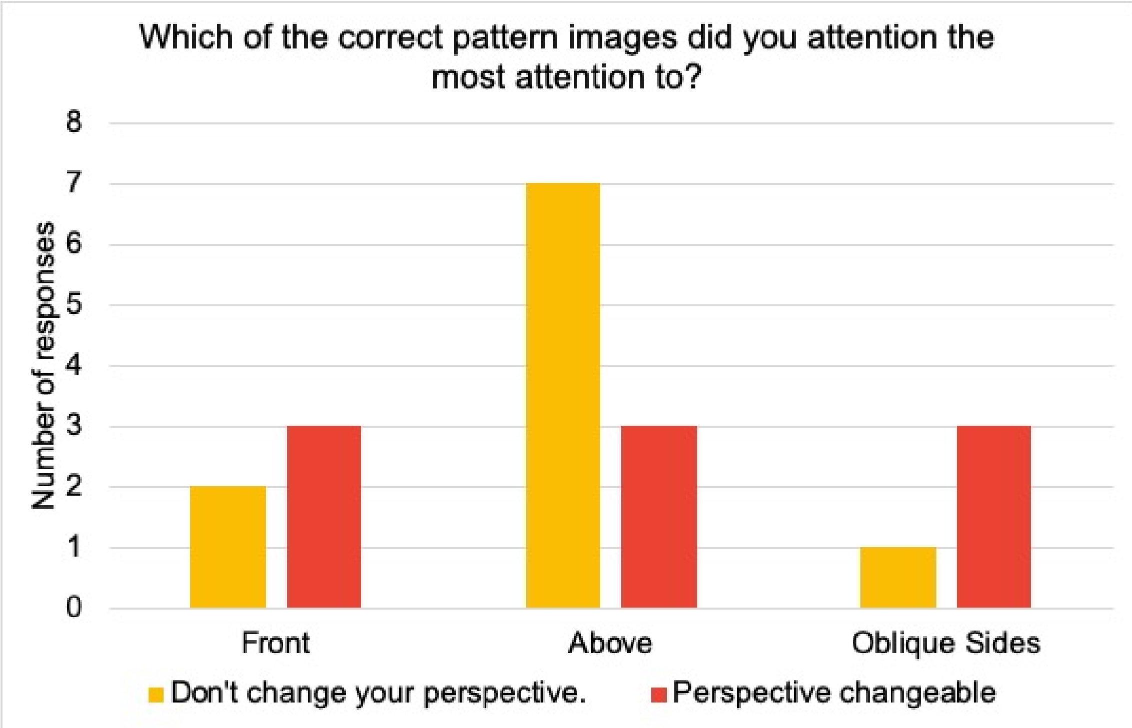

Figure 6.7: Which of the correct pattern images did you attention the most attention to?

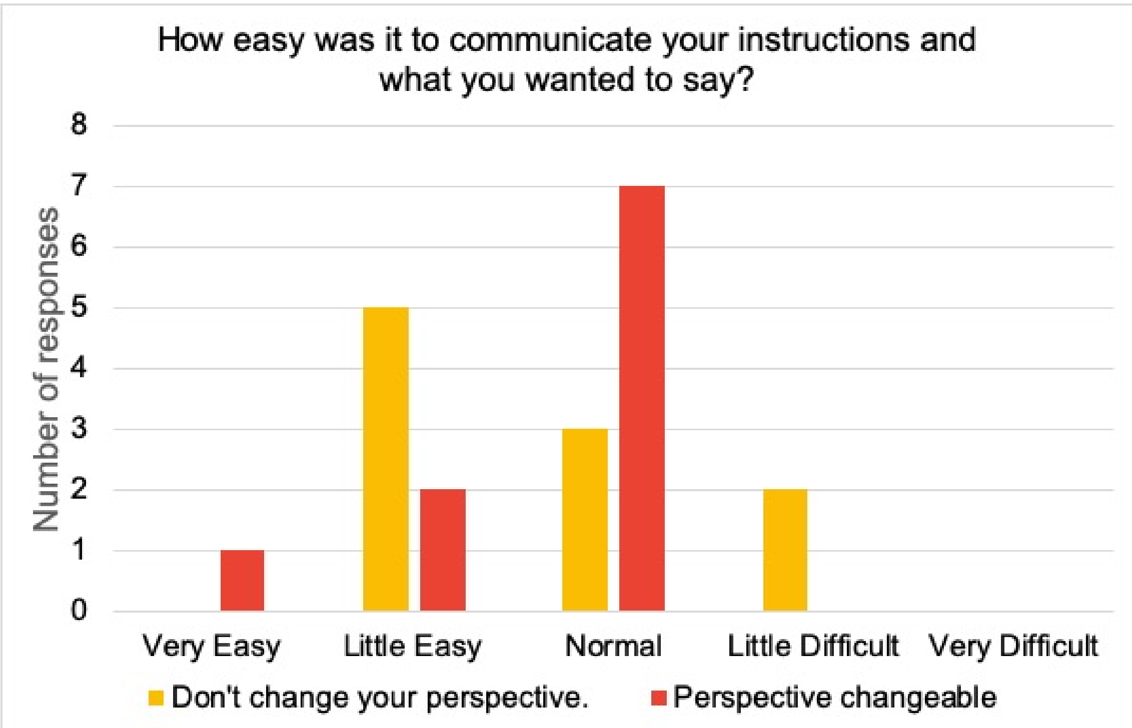

Figure 6.8: How easy was it to communicate your instructions and what you wanted to say?

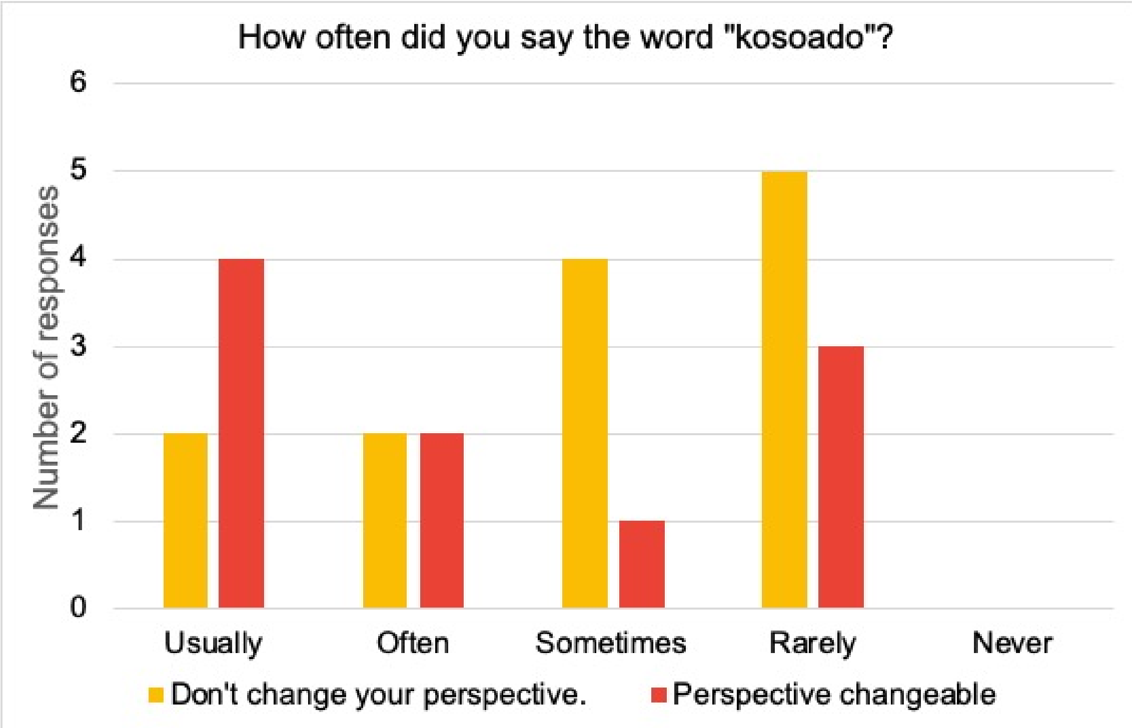

Figure 6.9: How often did you say the word "kosoado"?

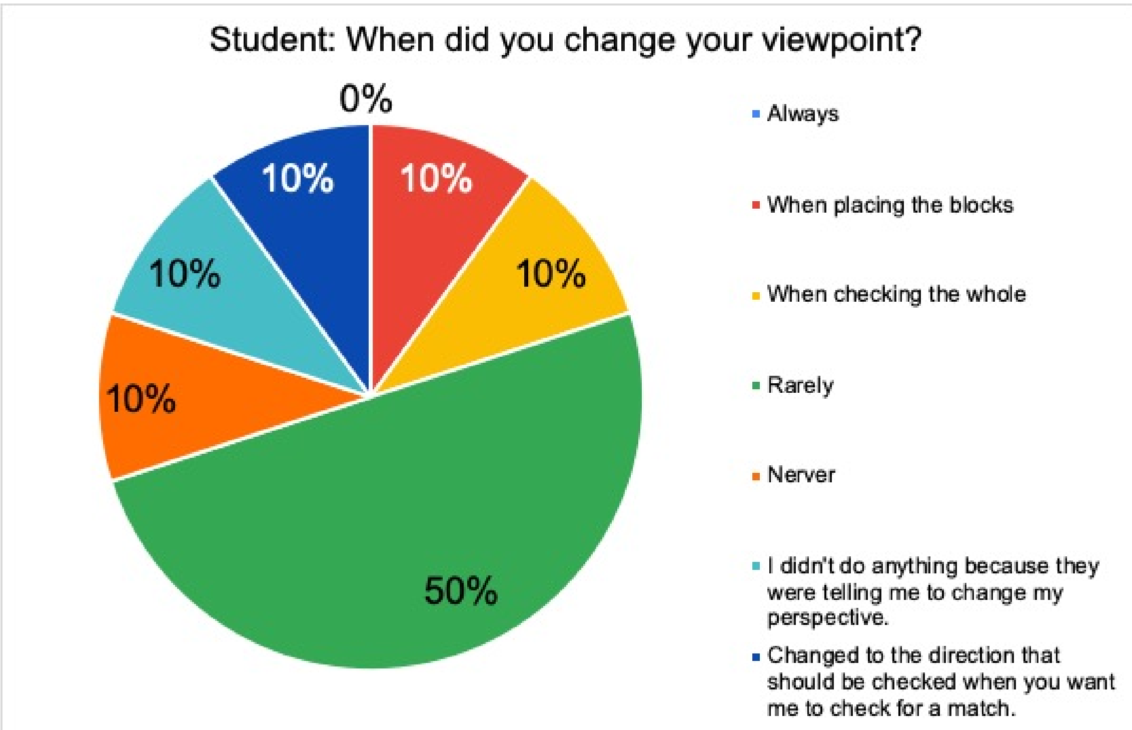

Figure 6.10: Student: When did you change your viewpoint?

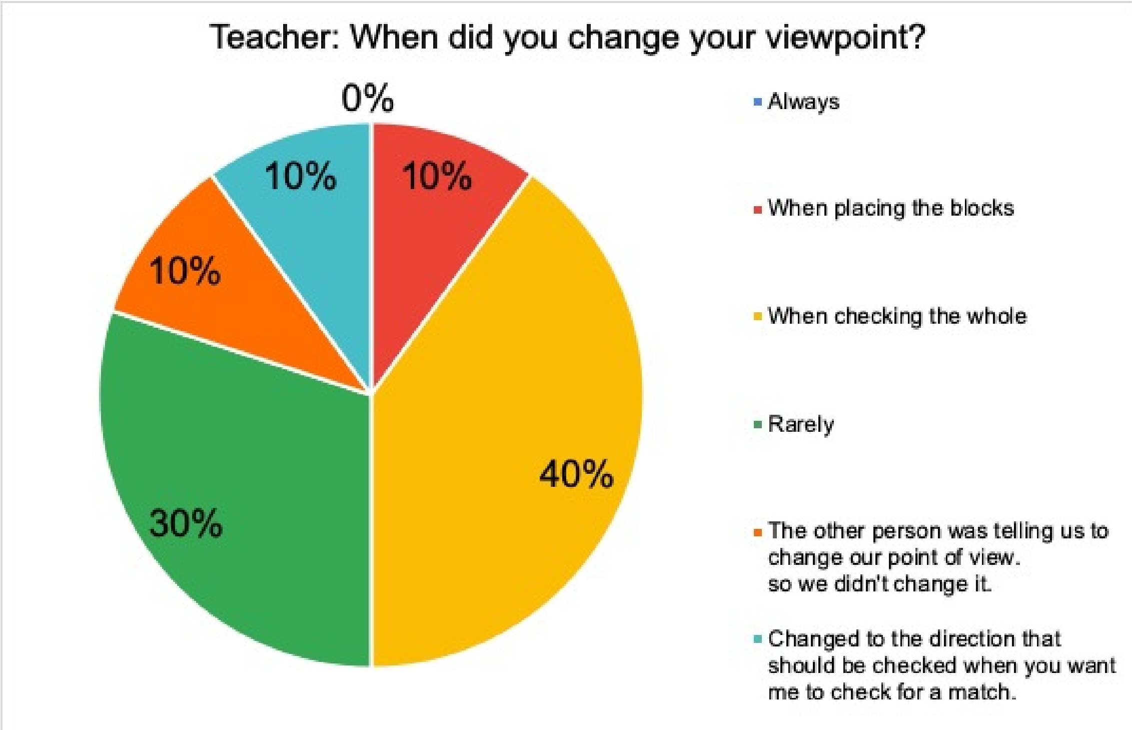

Figure 6.11: Teacher: When did you change your viewpoint?

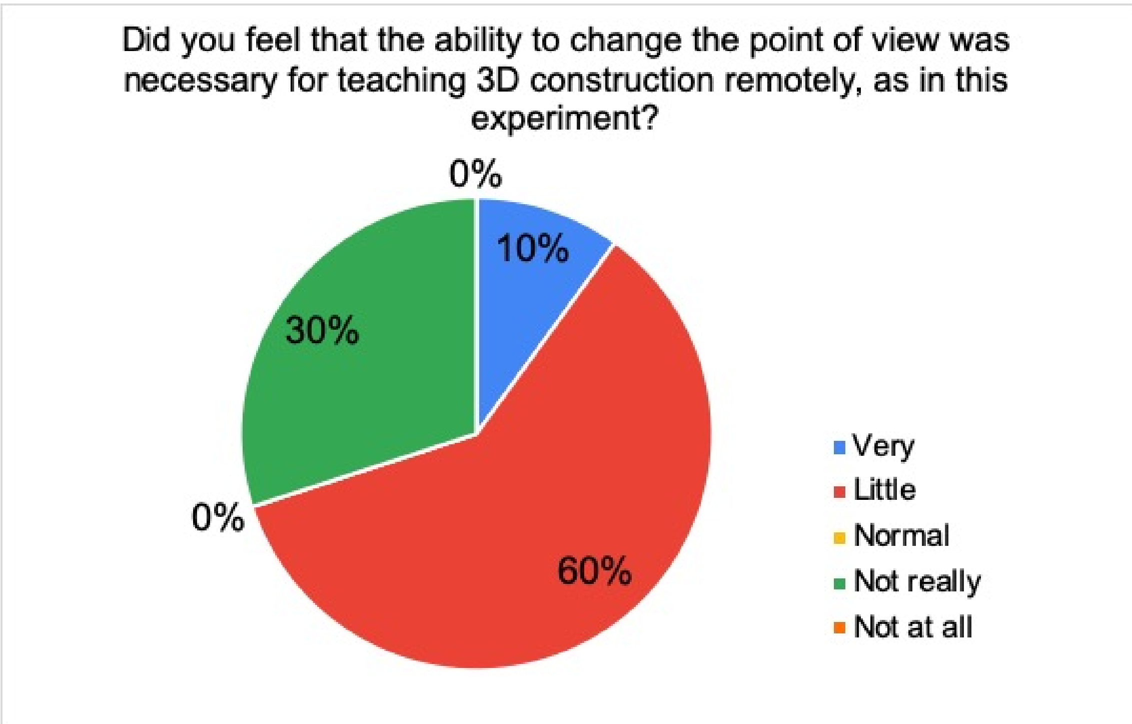

Figure 6.12: Did you feel that the ability to change the point of view was necessary for teaching 3D construction remotely, as in this experiment?

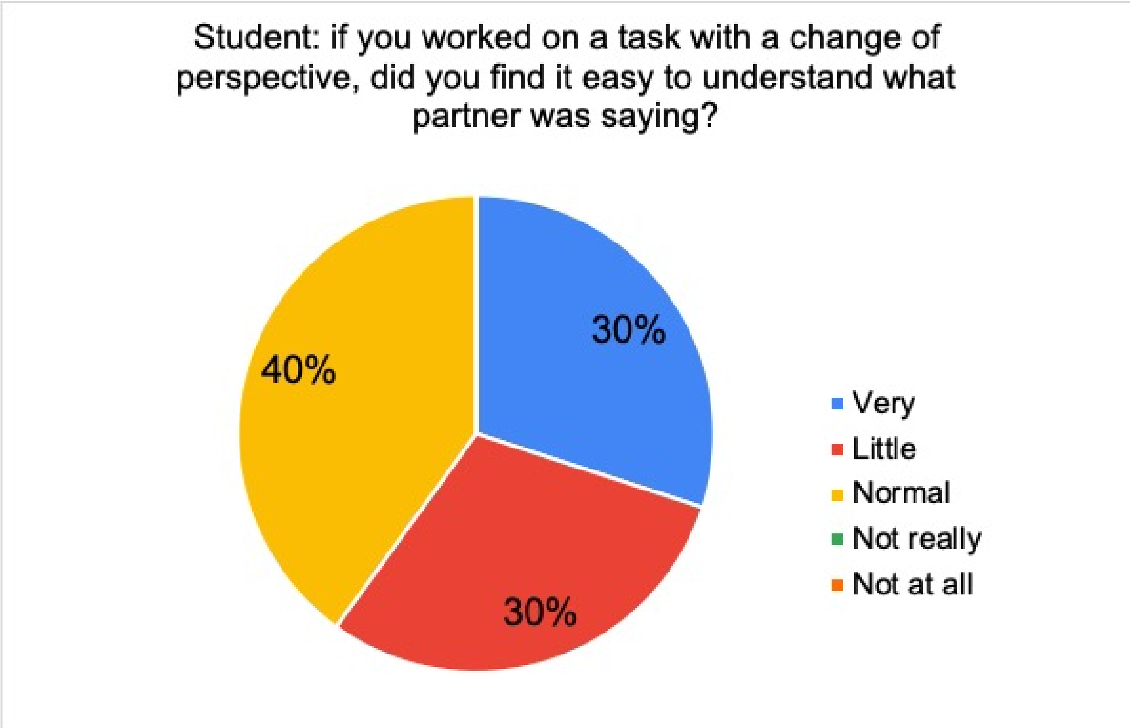

Figure 6.13: Student: If you worked on a task with a change of perspective, did you find it easy to understand what partner was saying?

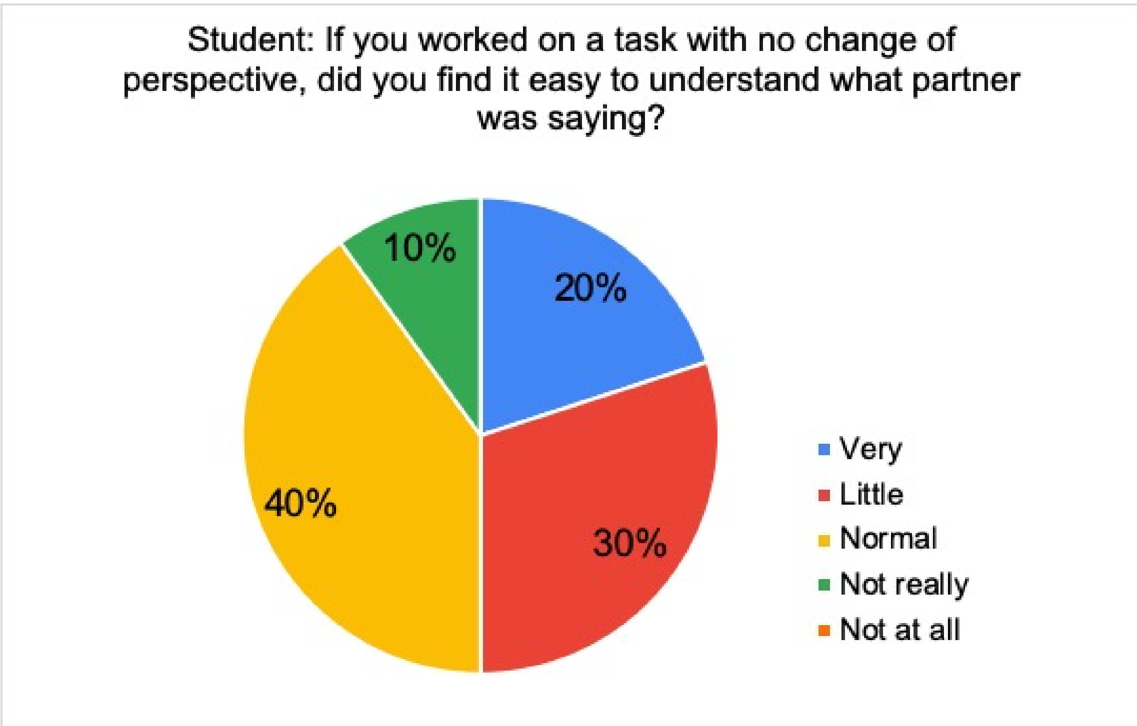

Figure 6.14: Student: If you worked on a task with no change of perspective, did you find it easy to understand what partner was saying?

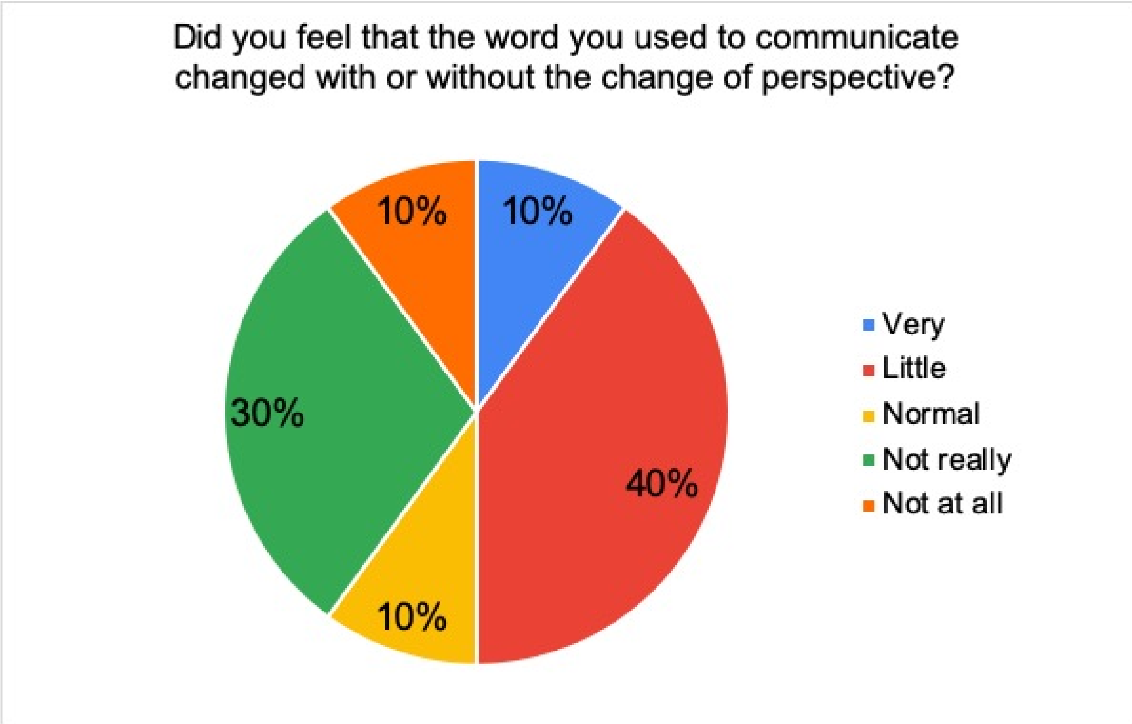

Figure 6.15: Did you feel that the word you used to communicate changed with or without the change of perspective?

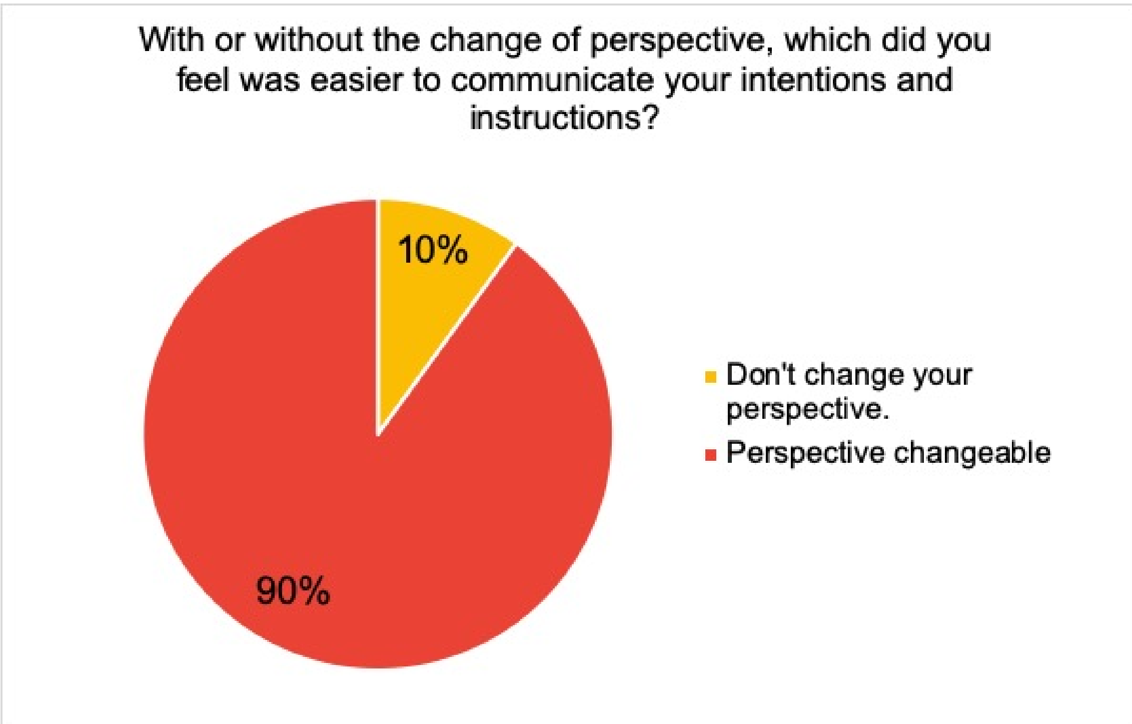

Figure 6.16: With or without the change of perspective, which did you feel was easier to communicate your intentions and instructions?

Chapter 7

Limitation and Discussion

7.1 Experiences of remote teaching and changes in communication methods for 3D objects

In Study 2, remote teaching was conducted on a Christmas tree. Participant 1 (P1) commented, "It was easy to understand what I was instructing because the viewpoint was changed when I was the one being taught and we were both able to explain the situation regarding depth. However, there were times when the decorations were not visible in the video, which was a bit of a problem". P3 commented, "I thought I could see the position of the pine cone (decoration) behind the drum (decoration), but I couldn't see the side of it, so it was difficult. According to these comments, the depth sensor method used in the iPad Pro was not able to read the Christmas tree correctly. The material used in Christmas trees is made of such a material that it reflects diffusely when the light hits it.

In Study 2, remote teaching was conducted for such difficult to measure objects, and the following positive comments were made about the experience of the system. P1 commented that "it was easy to talk about depth and to understand where the other person was indicating" and P2 commented that "the ability to change the viewpoint and zoom made it easier to teach". According to these comments, although there were parts that could not be read by the sensor correctly, the remote teaching itself was carried out and the person being taught was able to design as instructed. This suggests that people were able to use verbal communication as well as visual communication to supplement their description of the object's condition, thus enabling correct communication.

Accordingly, in Study 3, I conducted an experiment using building blocks with surface properties that were not difficult to measure with the ToF sensor. In their descriptions, the participants said that they could assemble the blocks if they listened to the teacher's voice reading out the instructions, and that they almost never changed their viewpoints. They also commented that they almost never changed their point of view, and that there were many areas where they could not see the sides even when they changed their point of view. These comments show that remote teaching is difficult even for the task of 3D construction when the method of construction, the method of assembly and the appearance of the assembled object are different. The results of this experiment suggest that, because the camera is placed in front of the building blocks, there are few surfaces that can be sensed in relation to the object, so it is difficult to see new surfaces in the images displayed in the 3D point cloud even if the viewpoint is changed. In addition, since the instructions for the design of the building blocks were arranged in order from the back, it was possible to assemble the blocks only by listening to the voice of the teacher. In other words, I showed that the need to know the depth data was not high in such use cases.

In this experiment(Chapter 6), based on the results of Study 3, I defined the design of the building blocks considering the motivation to move the viewpoint, abolished the instructions in order to investigate the change of communication during instruction with and without changing the viewpoint, and provided some practice time beforehand. When we asked the participants what they thought about and what words they used to communicate with their partner, most of them said that they communicated relative information such as the position and rotation angle of the blocks to the ones already placed. This may be because the building blocks are easy to explain and because there is always an adjacent surface somewhere. For example, the communication method may be different when the subject is not a remote lecture such as a fixed design, but a subject that needs to be followed and explained in real time, such as "Ayatori" or Japanese calligraphy.

In the future work, this study could be used to verify remote teaching, which requires real-time performance. As suggestions for improvement of the teaching method when the viewpoint change function was introduced to the videoconferencing tool, the following comments were made: "I thought it would be good if the correct image of the teacher could move the viewpoint in 3D", "I thought it would be helpful if the camera device itself could move". and "improving the accuracy and quality of the sensing images by using multiple cameras". Although these comments would add restrictions to the environment for remote teaching, the quality of the experience may be more important in terms of ease of communication.

7.2 Telecommunication Environment

With ToF sensing method, if the background is made of a material that diffusely reflects infrared light, the distance cannot be read correctly, resulting in a missing value and a black background 7.2. In this case, a non-reflective object can be placed as shown in Figure 7.2 (c). It is also possible to rotate a light weight object such as a Christmas tree while working on it to complete the decorating task.

In the proposed teleconference system, the video from the sensing device is displayed by the screen sharing function of the video conferencing tool, so a good communication environment is required. In Study 2 and 3, it was commented that the delay was too large. Therefore, in this experiment, we improved the communication environment and conducted the experiment. The discussion of the screen sharing method which is the cause of the delay is shown below.

There are three ways to import the video from iOS and iPad devices into ZOOM (Version 5.4.7).

1. Method to connect the iOS or iPad device directly to the PC via USB and select "iPhone/iPad via AirPlay" in "Share Screen" to display the video.
2. Method to connect an HDMI capture device from the iOS or iPad device to the PC via USB, and specify the device name in the camera selection.
3. Method to from an iOS or iPad device, connect to the USB port via an HDMI capture device, and select "Contents from 2nd Camera" under "Share screen" to display.

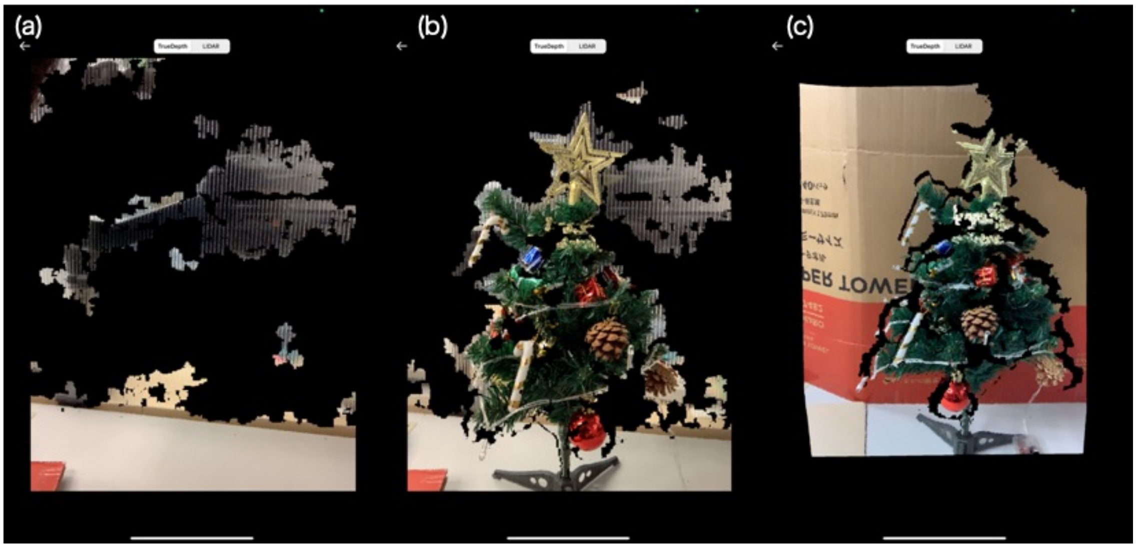

Figure 7.1: The background of the object: (a) A sensing view of transparent, rough-surfaced walls, (b) A sensing view of the Christmas tree placed in front of (a), (c) Improved view of the situation in (b)

Method 1 uses AirPlay to capture the screen, so the resolution of the shared screen is degraded and there is a large delay in the transmission of the video. Method 2 currently has the problem that the resolution of the shared image is fixed at 640x360, and "HD" in the video item of Zoom settings is ignored. Method 3 is said to be able to force the video to be delivered at the original resolution as long as there are CPU specs and a good network environment. Therefore, in this study, the video was displayed using method 3 to minimize the resolution degradation and delay of the shared screen as much as possible.

The displayed image would depend on the maximum possible output resolution of the video conferencing tool. The maximum resolution of zoom is 1920×1080 due to the screen sharing method used in this research. Also, the size of the screen that is drawn on the zoom screen depends on the size of the device because it is recognized as a camera device. In addition, screen sharing has a problem that the visible area is narrowed because the screen is shared in portrait mode while the zoom screen is in landscape mode.

Chapter 8

Conclusion

In this study, I proposed a remote communication system which introduced depth sensing technology. I also used it to conduct remote teaching on 3D built objects, and investigated the changes in people's communication. Until now, the number of individuals who own depth sensors has been small, and there has not been enough research to explore the use cases of depth sensing technology for personal use and to study the changes in communication using it. In this study, I focus on fields of telecommunication, starting from the consideration of the object to be photographed, and then I take up two specific use cases to investigate the possibility of new applications of depth sensors for personal users.

In this study, four experiments were conducted. In Study 1, a simple experimental design was used to investigate people's communication, as it was necessary to investigate in advance whether the relationship between the people in the video conference would lead to special behaviour. In order to verify this, I implemented a skeletal estimable system that works in real time in order to be able to analyse body movements from multiple camera images. As a result, it was judged that the relationship was not affected, the part of the screen where the movement was often confirmed was within the range where it was possible to operate at hand, and that targeting an object was a more preferable use of the sensor. In Study 2, I designed and implemented a telecommunication system using the depth sensor in a videoconferencing system. The system allows the user to change the viewpoint of the image of the caller in the video conference. I also adopted teaching as a task in which the results of communication are clearly visible in order to investigate the changes in people's communication. In Study 2, the teaching task was to decorate a Christmas tree. The results showed that the teaching itself was performed without any problem, although some parts of the Christmas tree were difficult to be detected because of the depth sensor system. The results also showed the interestingness of the teleteaching experience and the effectiveness of the proposed telecommunication system. In Study 3, I changed the experimental environment by changing the object to one that was easy to detect and by giving instructions to the teacher. The results showed that the participants were not motivated to change their viewpoints and that the proposed system was not necessary. Similarly, the results showed that remote teaching is difficult even for the task of 3D construction when the method of construction, the method of assembly and the appearance of the assembled object are different.

Finally, in the main experiment, I modified the experimental environment, such as the design of the building blocks and the placement of the camera, to investigate the change in communication with and without the change in viewpoint. As a result, in the survey questionnaire, 70% of the participants answered that the viewpoint change function is necessary, and 90% of the participants answered that the viewpoint change function makes it easier to communicate intentions and instructions, etc. In addition, the use of "kosoado" with and without the ability to change viewpoints was more common when the ability to change viewpoints was enabled. The results show that my proposed videoconferencing system with depth sensor is effective for remote teaching of Christmas tree decorating and building blocks assembly (depending on camera position and design).

In the future, it is necessary to investigate the use case of remote teaching, where real-time and depth data are more required to limit the relative instructions. The discussions and limitations of the system described in this paper will lead to further research into distance communication for personal use cases.

Acknowledgement

I would like to thank University of Tsukuba and Pixie Dust Technologies, Inc. for supporting.

I am also thankful to all the members of the Digital Nature Group at University of Tsukuba for discussion and feedback. I thank Jun-Li Lu, Keita Yamazaki, Manabu Nakano and Yoshiki Nagatani for their insightful comments.

References

[1] Hantong Xu, Jiamin Xu, and Weiwei Xu. Survey of 3d modeling using depth cameras. *Virtual Reality Intelligent Hardware*, 1(5):483 – 499, 2019. 3D Vision.

[2] Anupama K. Ingale and Divya Udayan J. Real-time 3d reconstruction techniques applied in dynamic scenes: A systematic literature review. *Computer Science Review*, 39:100338, 2021.

[3] Michael Zollhöfer, Matthias Nießner, Shahram Izadi, Christoph Rehmann, Christopher Zach, Matthew Fisher, Chenglei Wu, Andrew Fitzgibbon, Charles Loop, Christian Theobalt, and Marc Stamminger. Real-time non-rigid reconstruction using an rgb-d camera. *ACM Trans. Graph.*, 33(4), July 2014.

[4] Fangchuan Li, Shuangjia Liu, Luyu Jiang, Weihua Zhang, and Jin Zhou. Novel use of the intel realsense sr300 camera for foot 3d reconstruction. *Leather and Footwear Journal*, 20:145–152, 06 2020.

[5] A. Singandhupe and H. M. La. A review of slam techniques and security in autonomous driving. In *2019 Third IEEE International Conference on Robotic Computing (IRC)*, pages 602–607, 2019.

[6] Michael Bekele Maru, Donghwan Lee, Kassahun Demissie Tola, and Seunghee Park. Comparison of depth camera and terrestrial laser scanner in monitoring structural deflections. *Sensors*, 21(1):201, 2021.

[7] Lili Mu, Pantao Yao, Yuchen Zheng, Kai Chen, Fangfang Wang, and Nana Qi. Research on slam algorithm of mobile robot based on the fusion of 2d lidar and depth camera. *IEEE Access*, 8:157628–157642, 2020.

[8] V. Tadic, A. Odry, I. Kecskés, E. Burkus, Z. Király, and P. Odry. Application of intel realsense cameras for depth image generation in robotics. 2019.

[9] 2021-01-06. http://appft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=\%2Fnetahtml\%2FPTO\%2Fsearch-adv.html&r=1&f=G&l=50&d=PG01&p=1&S1=20100199228.PGNR.&OS=dn/20100199228&RS=DN/20100199228.

[10] 2021-01-06. https://patents.google.com/patent/US7852461B2/en.

[11] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. In *CVPR 2011*, pages 1297–1304, 2011.

[12] L. Yang, L. Zhang, H. Dong, A. Alelaiwi, and A. E. Saddik. Evaluating and improving the depth accuracy of kinect for windows v2. *IEEE Sensors Journal*, 15(8):4275–4285, 2015.

[13] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, and Andrew Fitzgibbon. Kinectfusion: Real-time 3d reconstruction and interaction using a moving depth camera. In *UIST '11 Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 559–568. ACM, October 2011.

[14] Zahoor Zafrulla, Helene Brashear, Thad Starner, Harley Hamilton, and Peter Presti. American sign language recognition with the kinect. In *Proceedings of the 13th International Conference on Multimodal Interfaces*, ICMI '11, page 279–286, New York, NY, USA, 2011. Association for Computing Machinery.

[15] Ke Sun, Wei Wang, Alex X. Liu, and Haipeng Dai. Depth aware finger tapping on virtual displays. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '18, page 283–295, New York, NY, USA, 2018. Association for Computing Machinery.

[16] Philipp Voigt, Matthias Budde, Erik Pescara, Manato Fujimoto, Keiichi Yasumoto, and Michael Beigl. Feasibility of human activity recognition using wearable depth cameras. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers*, ISWC '18, page 92–95, New York, NY, USA, 2018. Association for Computing Machinery.

[17] H. Park, J. Lee, and J. Ko. Achieving real-time sign language translation using a smartphone's true depth images. In *2020 International Conference on COMmunication Systems NETworkS (COMSNETS)*, pages 622–625, 2020.

[18] This latest camera sales chart shows the compact camera near death, 2021-01-06. https://petapixel.com/2017/03/03/latest-camera-sales-chart-reveals-death-compact-camera/.

[19] Sean Follmer, Daniel Leithinger, Alex Olwal, Akimitsu Hogge, and Hiroshi Ishii. Inform: Dynamic physical affordances and constraints through shape and object actuation. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, page 417–426, New York, NY, USA, 2013. Association for Computing Machinery.

[20] Y. Song, N. Zhou, Q. Sun, W. Gai, J. Liu, Y. Bian, S. Liu, L. Cui, and C. Yang. Mixed reality storytelling environments based on tangible user interface: Take origami as an example. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1167–1168, 2019.

[21] 2021-01-28. https://pdfpiw.uspto.gov/.piw?Docid=09829988.

[22] Yassin Kortli, Maher Jridi, Ayman Al Falou, and Mohamed Atri. Face recognition systems: A survey. *Sensors*, 20(2), 2020.

[23] Andreas Breitbarth, Timothy Schardt, Cosima Kind, Julia Brinkmann, Paul-Gerald Dittrich, and Gunther Notni. Measurement accuracy and dependence on external influences of the iPhone X TrueDepth sensor. In Maik Rosenberger, Paul-Gerald Dittrich, and Bernhard Zagar, editors, *Photonics and Education in Measurement Science 2019*, volume 11144, pages 27 – 33. International Society for Optics and Photonics, SPIE, 2019.

[24] Javier Alfaro-Santafé, Antonio Gómez-Bernal, Carla Lanuza-Cerzócimo, José Víctor Alfaro-Santafé, Aitor Pérez-Morcillo, and Alejandro Jesús Almenar-Arasanz. Three-axis measurements with a novel system for 3d plantar foot scanning: iphone x. *Footwear Science*, 12(2):123–131, 2020.

[25] Hang-Nga Mai and Du-Hyeong Lee. Accuracy of mobile device–compatible 3d scanners for facial digitization: Systematic review and meta-analysis. *J Med Internet Res*, 22(10):e22228, Oct 2020.

[26] Sujeet Mate, Igor D.D. Curcio, Ranjeeth Shetty, and Francesco Cricri. Mobile devices and professional equipment synergies for sport summary production. In *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*, TVX ’17 Adjunct, page 45–50, New York, NY, USA, 2017. Association for Computing Machinery.

[27] Roberto Yus, Eduardo Mena, Sergio Ilarri, Arantza Illarramendi, and Jorge Bernad. Multicamba: a system for selecting camera views in live broadcasting of sport events using a dynamic 3d model. *Multimedia Tools and Applications*, 74:4059–4090, 06 2015.

[28] Jeff Daemen, Jens Herder, Cornelius Koch, Philipp Ladwig, Roman Wiche, and Kai Wilgen. Semi-automatic camera and switcher control for live broadcast. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video*, TVX ’16, page 129–134, New York, NY, USA, 2016. Association for Computing Machinery.

[29] Moneish Kumar, Vineet Gandhi, Remi Ronfard, and Michael Gleicher. Zooming on all actors: Automatic focus+context split screen video generation. *Computer Graphics Forum*, 36, 04 2017.

[30] F. Daniyal and A. Cavallaro. Multi-camera scheduling for video production. In *2011 Conference for Visual Media Production*, pages 11–20, 2011.

[31] R. Panda and A. K. Roy-Chowdhury. Multi-view surveillance video summarization via joint embedding and sparse optimization. *IEEE Transactions on Multimedia*, 19(9):2010–2021, 2017.

[32] T. Koyama, I. Kitahara, and Y. Ohta. Live mixed-reality 3d video in soccer stadium. In *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.*, pages 178–186, 2003.

[33] Christian Lipski, Christian Linz, Kai Berger, Anita Sellent, and Marcus Magnor. Virtual video camera: Image-based viewpoint navigation through space and time. *Computer Graphics Forum*, 29(8):2555–2568, Dec 2010.

[34] Z. Cui, J. Gu, B. Shi, P. Tan, and J. Kautz. Polarimetric multi-view stereo. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 369–378, 2017.

[35] Hiroshi Sankoh, Sei Naito, Keisuke Nonaka, Houari Sabirin, and Jun Chen. Robust billboard-based, free-viewpoint video synthesis algorithm to overcome occlusions under challenging outdoor sport scenes. In *Proceedings of the 26th ACM International Conference on Multimedia*, MM '18, page 1724–1732, New York, NY, USA, 2018. Association for Computing Machinery.

[36] U. Yusuke, K. Takahashi, and T. Fujii. Free viewpoint video generation system using visual hull. In *2018 International Workshop on Advanced Image Technology (IWAIT)*, pages 1–4, 2018.

[37] Yanru Wang, Zhihao Huang, Hao Zhu, Wei Li, Xun Cao, and Ruigang Yang. Interactive free-viewpoint video generation. *Virtual Reality Intelligent Hardware*, 2(3):247 – 260, 2020. 3D Visual Processing and Reconstruction Special Issue.

[38] Janine Hacker, Jan vom Brocke, Joshua Handali, Markus Otto, and Johannes Schneider. Virtually in this together – how web-conferencing systems enabled a new virtual togetherness during the covid-19 crisis. *European Journal of Information Systems*, 29(5):563–584, 2020.

[39] Jiaxiong Hu, Qianyao Xu, Limin Paul Fu, and Yingqing Xu. Emojilization: An automated method for speech to emoji-labeled text. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19, page 1–6, New York, NY, USA, 2019. Association for Computing Machinery.

[40] Garreth W. Tigwell and David R. Flatla. Oh that's what you meant! reducing emoji misunderstanding. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, MobileHCI '16, page 859–866, New York, NY, USA, 2016. Association for Computing Machinery.

[41] Brid O'Conaill, Steve Whittaker, and Sylvia Wilbur. Conversations over video conferences: An evaluation of the spoken aspects of video-mediated communication. *Human–Computer Interaction*, 8(4):389–428, 1993.

[42] Marjolijn L. Antheunis, Alexander P. Schouten, and Joseph. B. Walther. The hyperpersonal effect in online dating: effects of text-based cmc vs. videoconferencing before meeting face-to-face. *Media Psychology*, 23(6):820–839, 2020.

[43] Cathy Cohen and Ciara R. Wigham. A comparative study of lexical word search in an audioconferencing and a videoconferencing condition. *Computer Assisted Language Learning*, 32(4):448–481, 2019.

[44] Zhengqing Li, Shio Miyafuji, Toshiki Sato, Hideaki Kuzuoka, and Hideki Koike. Omnieyeball: Spherical display equipped with omnidirectional camera and its application for 360-degree video communication. In *The 31st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings*, UIST '18 Adjunct, page 33–35, New York, NY, USA, 2018. Association for Computing Machinery.

[45] Zhengqing Li, Shio Miyafuji, Erwin Wu, Toshiki Sato, Hideaki Kuzuoka, and Hideki Koike. Omnieyeball: An interactive i/o device for 360-degree video communication. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces*, ISS '18, page 429–432, New York, NY, USA, 2018. Association for Computing Machinery.

[46] Kana Kushida and Hideyuki Nakanishi. A telepresence system using a flexible textile display. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, ICMI '16, page 412–413, New York, NY, USA, 2016. Association for Computing Machinery.

[47] Anthony DeVincenzi, Lining Yao, Hiroshi Ishii, and Ramesh Raskar. Kinected conference: Augmenting video imaging with calibrated depth and audio. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, CSCW '11, page 621–624, New York, NY, USA, 2011. Association for Computing Machinery.

[48] Sasa Junuzovic, Kori Inkpen, Tom Blank, and Anoop Gupta. Illumishare: Sharing any surface. In *CHI 2012*, pages 1919–1928. ACM, May 2012.

[49] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L. Davidson, Sameh Khamis, Mingsong Dou, Vladimir Tankovich, Charles Loop, Qin Cai, Philip A. Chou, Sarah Mennicken, Julien Valentin, Vivek Pradeep, Shenlong Wang, Sing Bing Kang, Pushmeet Kohli, Yuliya Lutchyn, Cem Keskin, and Shahram Izadi. Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, page 741–754, New York, NY, USA, 2016. Association for Computing Machinery.

[50] E. Saadatian, H. Samani, A. Vikram, R. Parsani, L. Tejada Rodriguez, and R. Nakatsu. Personalizable embodied telepresence system for remote interpersonal communication. In *2013 IEEE RO-MAN*, pages 226–231, 2013.

[51] Github - google-coral/project-posenet: Human pose detection on edgetpu, 2021-01-04. https://github.com/google-coral/project-posenet.

[52] Real-time human pose estimation in the browser with tensorflow.js | by tensorflow | tensorflow | medium, 2021-01-04. https://medium.com/tensorflow/real-time-human-pose-estimation-in-the-browser-with-tensorflow-js-7dd0bc881cd5.

[53] Usb accelerator | coral, 2021-02-03. https://coral.ai/products/accelerator/.

[54] Google cloud blog - news, features and announcements, 2021-01-04. https://cloud.google.com/blog/ja/products/gcp/what-makes-tpus-fine-tuned-for-deep-learning.

[55] tfjs-models/posenet at master · tensorflow/tfjs-models · github, 2021-01-04. https://github.com/tensorflow/tfjs-models/tree/master/posenet.

[56] Github - palakis/obs-websocket: Remote-control obs studio through websockets, 2021-01-04. https://github.com/Palakis/obs-websocket.

[57] Github - kirillmysnik/obs-ws-rc: asyncio-based python 3.5+ client to ‘obs-websocket‘ plugin for obs studio, 2021-01-04. https://github.com/KirillMysnik/obs-ws-rc.

[58] Us10268234b2 - bracket assembly for a multi-component vision system in an electronic device, 2021-01-04. https://patents.google.com/patent/US10268234B2/en.

[59] Us20170150124a1 - high resolution dot pattern, 2021-01-04. https://patents.google.com/patent/US20170150124A1/en.

[60] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR ’11, page 1297–1304, USA, 2011. IEEE Computer Society.

[61] Apple granted 49 patents today covering a 3d scanner/projector for a gaming console, an apple watch ui & more - patently apple, 2021-01-04. https://www.patentlyapple.com/patently-apple/2015/08/\newlineapple-granted-49-patents-today-covering-a-3d-scannerprojector-for-a-gaming\newline-console-an-apple-watch-ui-more.html.

[62] Apple advances projection system for possible future in-air gesturing feature for idevices & beyond - patently apple, 2021-01-04. https://www.patentlyapple.com/patently-apple/2016/01/apple-advances-projection-system-for-possible-future-in-air-gesturing-feature-for-\newlineidevices-beyond.html.

[63] Apple wins patent for a truedepth camera used to support hand gesturing in 3d interface for macs and beyond - patently apple, 2021-01-04. https://www.patentlyapple.com/patently-apple/2017/11/apple-wins-patent-for-a-truedepth-camera-used-to-support-hand-gesturing-in-3d-\newlineinterface-for-macs-and-beyond.html.

[64] While 3d facial recognition is coming to the iphone, a system for macs will also include presence technology - patently apple, 2021-01-04. https://www.patentlyapple.com/patently-apple/2017/07/while-3d-facial-recognition-is-coming-to-the-iphone-a-system-for-macs-will-also-include-\newlinepresence-technology.html.

[65] Apple wins a patent detailing the advancement of their lidar scanning system that could be included in the iphone 12 pro - patently apple, 2021-01-04. https://www.patentlyapple.com/patently-apple/2020/11/apple-wins-a-patent-detailing-the-advancement-of-their-lidar-scanning-system-that-\newlinecould-be-included-in-the-iphone-12-pro.html.

[66] Us10324171b2 - light detection and ranging sensor, 2021-01-04. https://patents.google.com/patent/US10324171B2/en.

[67] One of apple’s key inventions behind the ipad pro’s custom-designed lidar scanner was published today by uspto - patently apple, 2021-01-04. https://www.patentlyapple.com/patently-apple/2020/05/

one-of-apples-key-inventions-behind-the-ipad-pros-custom-designed-lidar-scanner-was-\newlinepublished-today-by-uspto.html.

[68] Apple wins a sophisticated lidar system patent covering the control of an autonomous vehicle in challenging conditions - patently apple, 2021-01-04. https://www.patentlyapple.com/patently-apple/2020/07/apple-wins-a-sophisticated-lidar-system-patent-covering-the-control-of-an-autonomous-\newlinevehicle-in-chal.html.

[69] Metal shading language specification, 2021-02-03. https://developer.apple.com/metal/Metal-Shading-Language-Specification.pdf.

[70] Streaming data from the truedepth camera - apple developer, 2021-01-04. https://developer.apple.com/documentation/avfoundation/cameras_and_media_capture/streaming_depth_data_from_the_truedepth_camera.

[71] Avfoundation | apple developer documentation, 2021-02-03. https://developer.apple.com/documentation/avfoundation.

[72] Arkit | apple developer documentation, 2021-02-03. https://developer.apple.com/documentation/arkit.

[73] 2021-01-04. https://developer.apple.com/videos/play/wwdc2018/503/.

[74] intrinsicmatrix | apple developer documentation, 2021-02-03. https://developer.apple.com/documentation/avfoundation/avcameracalibrationdata/2881135-intrinsicmatrix.

[75] 2021-01-04. https://developer.apple.com/documentation/multipeerconnectivity.

[76] 2021-01-04. https://developer.apple.com/documentation/modelio/mdlcamera/1391726-fieldofview.

Appendix

Pattern A 日本語版

1. まず「塗られていない長い直方体の積み木」を黒い線に縦に寝かせてください。
2. 同じ積み木を上に同じように重ねてください。
3. 青色の三角柱の積み木」を奥が高くなるように重ねてください。
4. 塗られていない長い直方体の積み木」を今置いている積み木の手前に左に縦に置いてください。
5. 同じ積み木をその左においてください。
6. 「短い円柱の積み木」を先程縦においた「塗られていない長い直方体の積み木」の上においてください。そのときに、上から見たときに円形に見えるように置いてください。
7. 同じ積み木を同様に、置いてください。
8. 「長い円柱の積み木」を２つの「短い円柱の積み木」の上の真ん中にまたがるようにおいてください。
9. 「塗られていない三角柱の積み木を」その上に、置いてください。そのときに、手前から見たら、三角形が見えて、左右対称になるようにおいてください。
10. 「緑色の厚さが薄い直方体の積み木」をその手前に寝かせて置いてください。
11. 同じ積み木をその上に同様に重ねて置いてください。
12. 「半円をくり抜いたような積み木」を先程の緑の上に面がぴったし重なるように置いてください
13. 「オレンジ色の半円柱の積み木」を２つ重ねて先程のくり抜いたような隙間にはめてください。
14. 「紫色の立方体の積み木」を４つ取り出し、緑色の積み木の手前に２個ずつ置いてください。
15. 「黄色の半円柱」を上から見たときに丸みを帯びるように置いてください。そのとき、手前から見たときに、手前から奥にアーチが作られるように配置してください。
16. 「半円をくり抜いたような積み木」を２つ取り出し、背中を合わせるようにして、「紫色の立方体の積み木」の手前に縦においてください。
17. 「短い円柱の積み木」をその上においてください。そのときに、上から見たときに円形に見えるように置いてください。

18. 同じ積み木を同様に置いてください。
19. 「オレンジ色の長い直方体の積み木」を「短い円柱の積み木」の横に「立方体の積み木」が１つおける空間を開けて右側においてください。
20. 同じ積み木をその上に同様に置いてください。
21. 「黄色の立方体の積み木」を「オレンジ色の長い直方体の積み木」の上に２つ横並びでおいてください。
22. 「オレンジ色の立方体の積み木」を「オレンジ色の長い直方体の積み木」の手前に２つ横並びに配置します。
23. 「緑色の立方体の積み木」を２つ、「オレンジ色の立方体の積み木」のそれぞれに重ねてください。
24. 「塗られていない三角柱の積み木」を「緑色の立方体の積み木」の上に左右対称になるようにかさねてください。
25. 「半円をくり抜いたような積み木」を「黄色の立方体の積み木」の上に手前から見たときにアーチになるように重ねてください。
26. 「黄色の厚さが薄い直方体の積み木」４つを一番広い面を重ねて、横方向に横長になるように隣接させて置いてください。このときに上から見たときに、１面のみ見えるように重ねてください。
27. 「青色の厚さが薄い直方体の積み木」２つを一番広い面を重ねて、縦方向にして、「半円をくり抜いたような積み木」に沿うように「黄色の厚さが薄い直方体の積み木」の上に置いてください。
28. 「長い円柱の積み木」をその右側に置いてください。そのときに、上から見たときに円形に見えるように置いてください。
29. 「緑色の三角柱の積み木」をその右に置きます。そのときに、手前から見たときに、三角形が見えて、斜辺が左に位置するように置いてください。
30. 「オレンジ色の三角柱の積み木」を「青色の厚さが薄い直方体の積み木」の上に置いてください。そのときに、手前から見たときに三角形が見えて、斜辺が左に位置するように置いてください。
31. 「青色の三角柱の積み木」を「オレンジ色の立方体の積み木」の右側に置いてください。そのときに、手前から見たときに三角形が見えて、斜辺が右に位置するように置いてください。
32. 同じ積み木を同じように手前に重ねて置いてください。
33. 「緑色の三角柱の積み木」を先程と同様に手前に２つ置いてください。
34. 「塗られていない立方体の積み木」を「青色の三角柱の積み木」の右側に置いてください。
35. 「緑色の三角柱の積み木」をその右側においてください。そのときに、手前から見たときに三角形が見えて、斜辺が右に位置するように置いてください。

36. その奥側に「青色の三角柱の積み木」を同様に並べて置いてください。
37. 「オレンジ色の三角柱の積み木」を「塗られていない立方体の積み木」の手前側に置いてください。そのときに、手前から見たときに、斜面が手前に見えるように置いてください。
38. 「オレンジ色の三角柱の積み木」を「青色の三角柱の積み木」の上においてください。そのときに、手前から見たときに、三角形が見えて、斜面が左に位置するようにおいてください。
39. 「黄色の三角柱」２つを背中が会うようにして、中央に配置した「半円をくり抜いたような積み木」の上に接地面があうように置いてください。
40. 完成

Pattern B 日本語版

1. 「黄色の厚さが薄い直方体の積み木」４つを一番広い面を重ねて、横方向に縦長になるように隣接させて黒色の線に沿わせるように手前に置いてください。
2. 「オレンジ色の直方体の積み木」をその右側に横方向に配置し、45 度手前方向に傾けて置いてください。そのときに、1 辺が隣接する積み木の辺に沿わせて置いてください。
3. 「塗られていない直方体の積み木」をその上に面が一致するように重ねて置いてください。
4. 「黄色の立方体の積み木」をその上の左側に重ねて置いてください。
5. 「オレンジ色の直方体の積み木」、「塗られていない直方体の積み木」、「紫色の立方体の積み木」の順番で下から高さ方向に重ねたものを先程と同様に左右対称に、置いてください。
6. 「塗られていない立方体の積み木」を右側の「オレンジ色の直方体の積み木」に隣接するように置いてください。
7. 「緑色の三角柱の積み木」をその上においてください。そのときに、手前から見たら、三角形の面が見えて、右側が斜面に位置するようにしてください。また、下の立方体の図形と面を一致させておいてください。
8. 「塗られていない立方体の積み木」を左右対称に左側の「オレンジの直方体の積み木」の隣に置いてください。
9. その上に、「青色の三角柱の積み木」を置いてください。そのときに、全体的に左右対称になっていることを確認してください。
10. 「紫色の立方体の積み木」を全体の右側の「塗られていない立方体の積み木」に面するように置いてください。
11. 「緑色の三角柱の積み木」をその上に同様の向きに置いてください。
12. 「紫色の立方体の積み木」を全体的に左右対称に位置する「塗られていない立方体の積み木」に面するように置いてください。

13. 「青色の三角柱の積み木」をその上に面が一致するようにおいてください。このときに全体的に左右対称になっていることを確認してください。

14. 「黄色の三角柱」2 つを左右対称に手前側にある「紫色の立方体の積み木」の手前に配置します。このときに面が一致するように置いてください。

15. 「青色の厚さが薄い直方体の積み木」２つを一番広い面を重ねて、中央の黄色積み木の上に寝かせて置きます。このときに、広い面が下になるようにおいてください。

16. 「半円をくり抜いたような積み木」を２つ横長に連結させて、その上に置いてください。そのときに、手前から見たときに、半円がくり抜かれているようにみえていることを確認してください。

17. 「黄色の半円柱の積み木」を右側の「半円をくり抜いたような積み木」の間に一致するように置いてください。

18. 「オレンジ色の半円柱の積み木」を左側に同様に配置してください。

19. 「緑色の三角柱」をその上に置いてください。そのときに、手前から見たときに直角となる角が上に来るように配置してください。

20. 「青色の三角柱」を同様に左右対称となるように配置してください。

21. 「緑色の厚さが薄い直方体の積み木」２つを左右対称に中央に４つ並んだ積み木の両端に奥行方向に配置してください。

22. 「半円をくり抜いたような積み木」をその上に門になるように重ねて置いてください。

23. 「オレンジ色の半円柱」をその上の右に、「黄色の半円柱」を左に重ねておきます。そのときに横から見たときに、半円になるように置いてください。

24. 「塗られていない直方体の積み木」２つを左右対称に中央の緑色の積み木の手前の外側の辺に沿うように縦長に置いてください。

25. 「半円をくり抜いたような積み木」を門になる向き中央の 2 つの「塗られていない直方体の積み木」の間に面が一致するように置いてください。

26. その上に「塗られていない三角形の積み木」を面が一致するように置いてください。

27. 「短い円柱の積み木」２つを左右対称に中央の「塗られていない直方体の積み木」の上に置いてください。そのとき、上から見たときに円形に見えることを確認してください。

28. 「オレンジ色の立方体の積み木」2 つを縦に面が一致するように重ねたものを、全体の右側の「紫色の立方体の積み木」の裏側に面が一致するように重ねてください。

29. 「黄色の立方体の積み木」をその上に面が一致するように重ねてください。

30. 「長い円柱」に「短い円柱」を面を一致させ重ねたものをその上においてください。

31. 「オレンジ色の三角柱」をその上においてください。そのときに、手前から見たときに左右対称に配置し、45 度内側に傾けておいてください。

32. 同様に左右対称に今のタワーを作ります。下から「緑色の立方体の積み木」2 つ、「紫色の立方体の積み木」、「長い円柱の積み木」、「短い円柱の積み木」、「オレンジ色の三角柱の積み木」の順番に重ねてください。それを左右対称の位置においてください。
33. 「緑の三角柱の積み木」を右側のタワーの「オレンジ色の立方体」2 つ分の面に沿うように配置してください。
34. 「青色の三角柱の積み木」も同様に左右対称に配置してください。
35. 最後に「塗られていない三角柱」を中央上に三角形の山に左右対称になるように重ねて下さい。
36. 完成

Pattern C 日本語版

1. 「塗られていない立方体の積み木」の縦の 1 辺を黒い線の真ん中に合わせて置いてください。
2. 「紫色の立方体の積み木」を左右対称に 2 つ面が一致するように手前側両サイドにおいてください。
3. 「紫色の立方体の積み木」２つを横にくっつけ、右側の「紫色の立方体の積み木」の右手前側の側面に置いてください。
4. 左右対称に「オレンジ色の立方体の積み木」２つを横にくっつけ、左側も同様に側面に左右対称になるように置いてください。
5. 「緑色の厚さが薄い直方体の積み木」２つを広い面を重ねて縦に、ちょうど真ん中の角に置いてください。そのときに、上から見たときに、2 つの積み木の接地面が左上から右下に見えていることを確認してください。
6. 「塗られていない立方体の積み木」をその上に面が一致するように重ねてください。
7. 「塗られていない直方体の積み木」2 つを手前から 2 つ目の積み木の受けに面を完全一致させ置いてください。
8. 「短い円柱の積み木」２つをそれぞれその上に重ねてください。
9. 「黄色の半円柱の積み木」２つをそれぞれその上に置いてください。そのときに、真横から見たときに、半円になっていることを確認してください。
10. 「塗られていない直方体の積み木」2 つを左右対称にそれぞれ、「オレンジ色の立方体の積み木」２つの奥側の側面と「紫色の立方体の積み木」２つの奥側の側面に沿わせて置いてください。
11. 「緑色の立方体の積み木」２つを横に繋げ、「紫色の立方体の積み木」の内側と緑色の積み木の間に置いてください。
12. 「黄色の立方体の積み木」２つも同様に左右対称になるように置いてください。
13. 「オレンジ色の直方体の積み木」２つを横に繋げ、手前の隙間に配置してください。そのときに、上から見たら、接地面が左上から右下に掛けてみえることを確認してください。

14. 「黄色の厚さが薄い直方体の積み木」２つをそれぞれ両サイドの側面に奥から１パーツ手前の立方体から沿わすように置いてください。

15. 「半円をくり抜いたような積み木」に「オレンジ色の半円柱の積み木」を一致するようにいれた積み木２つをそれぞれの両サイドの側面に奥から沿わすように置いてください。

16. 「塗られていない三角形の積み木」を２つそれぞれ、「黄色の厚さが薄い直方体の積み木」と「半円をくり抜いたような積み木」どちらもにも面するように置いてください。そのとき、手前から見たら、長方形がみえるように配置します。

17. 「オレンジ色の三角柱の積み木」２つを上から見たときに左右対称のリボンになるように並べ、中央手前の「オレンジ色の直方体の積み木」の両サイドに中心よりに置きます、

18. 「黄色の三角柱の積み木」をその間に置きます。

19. 「半円をくり抜いたような積み木」２つを真ん中が正円になるように繋げ、その手前に置いてください。そのときに、上から見たら、正円になるように置いてください。

20. 「短い円柱の積み木」をその正円に一致するように置いてください。

21. 「青色の厚さが薄い直方体の積み木」２つを辺が長くて、狭い面を横方向に重ねて、「塗られていない三角形の積み木」の手前側に置いてください。そのときに、広い面がそれぞれ下に接するように手前に置いてください。また、中央の図形に側面が沿うように置いてください。

22. 「黄色の厚さが薄い直方体の積み木」も左右対称になるように同様に配置してください。

23. 「緑色の三角柱の積み木」２つを一番広い面同士を背中合わせになるように重ねたものを、もう１ペアつくり、「青色の厚さが薄い直方体の積み木」の奥と手前の外側にそれぞれ配置してください。

24. 「青色の厚さが薄い直方体の積み木」も同様に左右対称になるように置いてください。

25. 「短い円柱の積み木」を中央の「塗られていない立方体の積み木」の上に面が合わさるようにおいてください。

26. 「青色の三角柱の積み木」をその上に置いてください。そのときに、手前から見たときに、三角形がみえるように置いてください。

27. 最後に「長い円柱の積み木」２つを横向きに中心に置いてください。そのときに、積んでいる積み木に沿うように置いてください。

28. 完成

Pattern D 日本語版

1. ボックスの左にある「半円をくり抜いたような積み木」２つと「黄色の半円の積み木」２つからなる塊を奥行方向に黒い線に置いてください。

2. 「長い円柱の積み木」をその上の奥側に置いてください。

3. 「短い円柱の積み木」をその上に置いてください。

4. 「塗られていない立方体の積み木」２つを縦に重ねて、手前の面に一致するように置いてください。

5. 「塗られていない直方体の積み木」２つを手前の積み木の両サイドの面に一致するように置いてください。

6. 「オレンジ色の直方体の積み木」２つを左右対称に手前の積み木の前に立てて置いてください。

7. 「緑色の三角柱の積み木」２つをそれぞれの寝かせてオレンジ色の積み木に正方形の面が接するように置いてください。このときに、真横から見たときに、三角形になっていることを確認してください。

8. 「青色の厚さが薄い直方体の積み木」２つを、両端にそれぞれ奥行方向に置いてください。

9. 「黄色の厚さが薄い直方体の積み木」４つの面積が一番大きい面で重ねたものを奥側に横方向に置いてください。

10. 「青色の三角柱の積み木」を縦方向に黄色の積み木の両端においてください。そのとき、斜面が内側になるようにおいてください。

11. 「紫色の立方体の積み木」を右側の「青色の厚さが薄い直方体の積み木」の上に奥行き方向に詰めておいてください。

12. 「緑の立方体の積み木」を同様に左右対称においてください。

13. 「半円をくり抜いたような積み木」２つそれぞれをそれらの立方体の上においてください。そのときに、正面から見たときに、半円をくり抜いたように見えることを確認ください。

14. 「短い円柱」２つをその半円に一致するように置いてください。

15. 「黄色の三角形の積み木」２つをそれぞれ、「青色の厚さが薄い直方体の積み木」の上に置きます。そのときに、一番広い面が奥側の側面に接するように置いてください。1

16. 「緑色の三角柱の積み木」をオレンジ色の積み木の両端に置いてください。そのとき、正方形の面がオレンジ色の面に接するようにおいてください。

17. ４つの立方体を使って正方形の面を持つ直方体を作ります。左上が「緑色」、右上が「オレンジ色」、左下が「紫色」、右下が「黄色」です。それらの積み木を組み合わせたら、左の奥の積み木の横に置いて、奥行方向に 45 度傾けてください。そのとき、奥の縦の辺が接するように回転してください。

18. 「青色の三角柱の積み木」をその上においてください。そのときに、垂直方向に直角にならないように置いてください。

19. 右側も、同様に４つの立方体を使って正方形の面を持つ直方体を作ります。左上が「黄色」、右上が「紫色」、左下が「オレンジ色」、右下が「紫色」です。それらの積み木を組み合わせたら、右の奥の積み木の横に置いて、奥行方向に 45 度傾けてください。そのとき、奥の縦の辺が接するように回転してください。

20. 「青色の三角柱の積み木」をその上に上においてください。そのときに、左右対称になるようにおいてください。
21. 「緑色の厚さが薄い直方体の積み木」を紫色の２つの面に接するように奥に置きます。そのとき、広い面が下に位置するように置きます。
22. 「オレンジ色の半円柱」をその上の手前に置きます。そのとき、横から見たときに、半円がみえるように置きます。
23. 「オレンジ色の三角柱」をその隣の奥側に置きます。そのときに、横から見たときに、三角形が見えるように置きます。
24. 全体の逆側も同様に左右対称におきます。
25. 「長い円柱の積み木」を手前の右の２つの緑色の三角柱の間に立てておきます。
26. 「短い円柱の積み木」をその上におきます。
27. 「塗られていない三角柱の積み木」を山が中央上に来るように置きます。
28. 完成

Pattern A English Version

1. Lay a long unpainted rectangular block vertically on the black line.
2. Put the same blocks on top of each other in the same way.
3. Put the blue triangular blocks on top of each other so that the back is higher.
4. Place a long unpainted rectangular block vertically to the left of the current block.
5. place the same block to the left of it.
6. Place the short cylinder block on top of the long unpainted rectangular block you just placed vertically. Place the short cylinder blocks on top of the long unpainted rectangular blocks, so that they look circular when viewed from above. 7.
7. Place the same blocks in the same way.
8. Place the long block of wood across the middle of the two short blocks of wood.
9. Place the unpainted triangular blocks on top. Place the unpainted triangular blocks on top of the painted triangular blocks so that the triangles are visible and symmetrical when viewed from the front.
10. Place a green, thin, rectangular block of wood in front of it.
11. Put the same blocks on top of each other in the same way.
12. Place the semi-circular blocks on top of the green blocks so that their faces are exactly on top of each other.
13. Place two orange semi-circular blocks on top of each other in the hollowed-out space.
14. Take the four purple cubes and place two in front of each of the green cubes.

15. Place the yellow half-cylinders so that they are rounded when viewed from above. Place the yellow semi-circular columns so that when viewed from the front they form an arch from the front to the back.

16. Take the two half-circle blocks and place them back to back, vertically in front of the purple cube blocks.

17. Put the short cylinder blocks on top of them. Place the short cylinder block on top of the short cylinder block so that it looks like a circle when viewed from above.

18. Place the same blocks in the same way.

19. Place the long orange rectangular block on the right side of the short cylinder block, leaving room for one cube block.

20. Place the same blocks on top of each other in the same way.

21. Place two yellow cubes side by side on top of a long orange rectangular block.

22. Place two orange cubes side by side in front of the orange long rectangular blocks.

23. Place two green cubes on top of each of the orange cubes.

24. Stack the unpainted triangular blocks symmetrically on top of the green cubes.

25. Hollowed out half circles on top of yellow cubes in an arch.

26. Place the four yellow thin rectangular blocks next to each other with the widest side on top of each other, so that they form a long horizontal line. The blocks should be stacked so that only one side is visible when viewed from above.

27. Place the two blue thin rectangular blocks on top of the yellow thin rectangular blocks, with the widest side on top of each other, lengthwise, along the semi-circular blocks. Place the long column block on its right side. Place the long cylinder on its right side so that it looks like a circle when viewed from above.

28. Put the green triangular blocks on the right of the green triangular blocks. Place the green triangular block to the right of the green triangular block, so that when viewed from the front, the triangle is visible and the diagonal is to the left.

29. Place the orange triangular block on top of the blue thin rectangular block. Place the orange triangular blocks on top of the blue thin rectangular blocks so that the triangle is visible from the front and the diagonal is on the left.

30. Place the blue triangular blocks on the right side of the orange cubes. Place the blue triangular block on the right side of the orange cube block so that the triangle is visible from the front and the diagonal is on the right.

31. Place the same blocks on top of each other in the same way.

32. Place two green triangular blocks in front of each other in the same way as before.

33. Place the unpainted cubes to the right of the blue triangular blocks.

34. Place the green triangular block to the right of the green triangular block. Place the green triangular block on the right side of the green triangular block so that the triangle is visible from the front and the diagonal is on the right. Place the blue triangular building blocks on the far side of the triangular building blocks.
35. Place the orange triangular blocks in front of the unpainted cubes. Place the orange triangular blocks in front of the unpainted cubes so that the slope is visible from the front.
36. Place the orange triangular blocks on top of the blue triangular blocks. Place the orange triangular blocks on top of the blue triangular blocks so that the triangle is visible from the front and the slope is on the left.
37. Place the two yellow triangles back to back on top of the semi-circular building blocks in the centre so that they meet on the ground.
38. Complete

Pattern B English Version

1. Place the four yellow thin rectangular blocks in front of each other with their widest sides facing each other, horizontally and vertically, along the black line.
2. Place the orange rectangular blocks horizontally to the right of the orange rectangular blocks, tilting them 45 degrees towards you.
3. Place the unpainted rectangular blocks on top of the painted rectangular blocks so that their faces are aligned.
4. Place the yellow cubes on top of the yellow cubes on the left side
5. Place the orange rectangular blocks, the unpainted rectangular blocks and the purple cubic blocks in the same symmetrical order, stacked from bottom to top
6. Place the unpainted cubes next to the orange cubes on the right.
7. Place the green triangular block on top of it. The green triangular block should be placed on top of the green triangular block so that the triangular face is visible from the front and the right side is on the slope. Make sure that the faces are aligned with the cubes below. 8.
8. Place the unpainted cube symmetrically next to the orange rectangular block on the left.
9. Place a blue triangular block on top of it. Make sure there is symmetry throughout. 10.
10. Place the purple cube block on top of the unpainted cube block on the right side of the whole.
11. Place the green triangular blocks on top of the green triangular blocks in the same direction.

12. Place the purple cubes facing the unpainted cubes in an overall symmetrical position.
13. Place the blue triangular blocks on top of the unpainted cubes so that their faces coincide. Make sure there is symmetry throughout. 14.
14. Place the two yellow triangular prisms symmetrically in front of the purple cube blocks. Make sure that the faces are aligned.
15. Place the two blue thin rectangular blocks on top of each other with the widest side on top of the central yellow block. The wide side should be on the bottom. 16.
16. Place the two semi-circular blocks on top of each other in a horizontal pattern. Make sure they look like hollowed out half circles when viewed from the front.
17. Place the yellow semi-circular blocks in line with the hollowed-out semi-circle blocks on the right.
18. Place the orange semi-circular blocks in the same way on the left.
19. Place the green triangular blocks on top. Place the green triangle on top of the green triangle so that the right angle is on top of the triangle when viewed from the front. 20.
20. Place the blue triangles in the same symmetrical way.
21. 2 green thin rectangular blocks must be placed symmetrically at each end of the 4 central blocks.
22. "Hollowed out half-circle blocks" are placed on top of each other to form a gate.
23. Place the orange semi-circular blocks on the right and the yellow semi-circular blocks on the left. Place the orange half-circle on the right and the yellow half-circle on the left so that they form a semicircle when viewed from the side.
24. Place the two unpainted rectangular blocks symmetrically along the outer edge of the central green block.
25. Place the "hollowed out half-circle blocks" face up between the two unpainted rectangular blocks facing the gate.
26. Place an unpainted triangular block of wood on top of the unpainted triangular block of wood, so that the faces are congruent.
27. Place the two short cylinder blocks symmetrically on top of the central unpainted rectangular block. Make sure that they look circular when viewed from above.
28. Place the two orange cubes, face to face, vertically on top of each other and face to face on the back of the purple cubes on the right side of the whole.
29. yellow cubes on top of each other, face to face.
30. Place a long cylinder on top of a short cylinder with matching faces. Place the orange triangle on top of the long cylinder. Place the orange triangular pillar on top of the orange triangular pillar in a symmetrical position, tilted inwards by 45 degrees. 32.

31. Build the current tower in the same symmetrical way. From the bottom, place two green cubes, a purple cube, a long cylinder, a short cylinder and an orange triangular cylinder. Place them in a symmetrical position. 33.
32. Place the green triangular blocks along the faces of the two orange cubes on the right-hand tower.

Pattern C English Version

1. Place the unpainted cubes with one vertical side in the middle of the black line.
2. Place two symmetrical purple cubes on either side of the foreground so that their sides are aligned.
3. Place the two purple cubes side-by-side on the right front side of the purple cubes.
4. Symmetrically place the two orange cubes side-by-side on the left side, also symmetrically on the right side.
5. Place the two green thin rectangular blocks vertically with their wide sides overlapping, right in the middle corner. Make sure that when viewed from above, the ground surface of the two blocks is visible from top left to bottom right.
6. Place an unpainted cube on top of the unpainted cube so that the faces are aligned.
7. Place the two unpainted rectangular blocks on top of the second block from the front with their faces perfectly aligned.
8. Place two short cylinder blocks on top of each other.
9. Place two yellow half-cylinder blocks on top of each other. Make sure they form a semicircle when you look at them from the side. 10.
10. Place the two unpainted rectangular blocks symmetrically along the far side of the two orange cubes and along the far side of the two purple cubes.
11. join the two green cubes horizontally and place them between the inside of the purple cubes and the green cubes.
12. Place the two yellow cubes in the same symmetrical way.
13. Connect the two orange rectangular blocks horizontally and place them in the gap in front of you. Make sure that when you look at them from above, the ground surface is from the top left to the bottom right.
14. Place two yellow thin rectangular blocks on each side, starting from the cube one part in front of the other.
15. Place two matching orange half-cylinder blocks in a hollowed-out half-circle, along each side from the back.
16. Place the two unpainted triangular blocks so that they face both the thin yellow rectangular blocks and the hollowed-out half-circle blocks. The rectangles should look like rectangles when viewed from the front.

17. Place the two orange triangular blocks in a symmetrical ribbon when viewed from above, on either side of the orange rectangular block in the centre.
18. The yellow triangular blocks are placed between the two. The yellow triangular blocks are placed between them.
19. Place them in front of each other so that they form a perfect circle when viewed from above.
20. Place the short cylinder block in line with the circle.
21. Place two blue thin rectangular blocks with long sides and narrow sides on top of each other in front of the unpainted triangular blocks. Place them in front of the unpainted triangular blocks, so that the wider sides touch each other below. The sides should be in line with the central figure.
22. Yellow thin rectangular blocks must be placed in the same way to make them symmetrical.
23. Make another pair of green triangular blocks with the widest sides back to back and place them on the outside of the blue thin rectangular blocks.
24. Place the blue thin rectangular blocks in a symmetrical way.
25. Place the short cylinder blocks on top of the unpainted cube blocks in the centre, so that their faces meet.
26. Place the blue triangular blocks on top. Place the blue triangular blocks on top of the blue triangular blocks so that the triangles are visible when viewed from the front. 27.
27. Finally, place the two long cylinder blocks horizontally in the centre. Place them horizontally in the centre so that they are in line with the other blocks in the stack.
28. finish

Pattern D English Version

1. Place the block consisting of the two hollowed-out half-circle blocks and the two yellow half-circle blocks on the black line at the left of the box in the direction of the back.
2. Place the long blocks on the far side of the black line
3. Place a short block of cylinders on top of it
4. Place the two unpainted cubes vertically on top of each other so that they are on the same side
5. Place two unpainted rectangular blocks on top of each other so that they are on the same side of the first block
6. Two orange rectangular blocks must be placed symmetrically in front of the first block
7. Place the two green triangular blocks on their sides so that their square faces touch the orange blocks. Make sure they form a triangle when you look at them from the side.

8. Place two blue thin rectangular blocks at each end of the puzzle, one at the back and one at the front.
9. 4 yellow thin rectangular blocks, stacked on top of each other on the largest side, placed horizontally at the back.
10. Place the blue triangular blocks vertically at the ends of the yellow blocks. Place the blue triangular blocks vertically at each end of the yellow blocks, with the slopes facing inwards.
11. Place the purple cubes on top of the blue thin rectangular blocks on the right, in the direction of depth.
12. Place the green cubes symmetrically in the same way.
13. Place each of the two semi-circular blocks on top of the cubes. Make sure they look like a hollowed-out half-circle when viewed from the front.
14. Place the two "short columns" on top of each other so that they coincide with the semicircle.
15. Place each of the two yellow triangular blocks on top of the blue thin rectangular block. Place the two yellow triangular blocks on top of each blue thin rectangular block, with the widest side touching the far side.
16. Place the green triangular blocks on either side of the orange blocks. Place the green triangular blocks on either side of the orange blocks, with the square side touching the orange side.
17. Use the four cubes to make a rectangle with a square face. The top left is 'green', the top right is 'orange', the bottom left is 'purple' and the bottom right is 'yellow'. When you have put these blocks together, place them next to the back left block and tilt them 45 degrees towards the back. The blocks should be rotated so that the vertical sides of the back blocks touch.
18. Place the blue triangular block on top of the blue triangular block. Place the blue triangular block on top of the blue triangular block, making sure that it is not perpendicular to the vertical axis.
19. On the right side, use the same four cubes to make a rectangle with a square face. The top left is 'yellow', the top right is 'purple', the bottom left is 'orange' and the bottom right is 'purple'. When you have put these blocks together, place them next to the right-hand back block and tilt them 45 degrees towards the back. The blocks should be rotated so that the vertical sides of the back blocks touch.
20. Place the blue triangular block on top of the blue triangular block. Place the blue triangular blocks on top of the blue triangular blocks so that they are symmetrical.
21. Place the green thin rectangular blocks at the back so that they touch the two purple sides. The wide side should be at the bottom.

22. Place the orange half-cylinder in front of the top. Place the orange semi-circle in front of it so that the semi-circle is visible when viewed from the side.
23. Place the orange triangular pillar on the far side next to it. The orange triangle is placed on the far side so that the triangle is visible when viewed from the side.
24. Place the other side of the whole symmetrically.
25. Place the long column block between the two green triangles on the right.
26. Place the short blocks on top of the long blocks.
27. Place the unpainted triangular blocks so that the pile is in the top centre.
28. Complete