International Journal of Service and Knowledge Management
International Institute of Applied Informatics
2023, Vol. 7, No. 1, IJSKM737

Developing an AR Lecture Recording System with Direct Manipulation of Virtual Slides by Physical Objects

Yuma Ito *, Masato Kikuchi *, Tadachika Ozono *

Abstract

The increasing demand for lecture videos needs support for making them. Some lecturers conduct conventional and physical-style presentations in the real world. Videos recording slides physically displayed by a projector are blurrier than videos recording slides on a display monitor. We propose a recording system that merges physical-style lectures and slides with electronic presentation tools. Hence, we aim to develop a lecture recording system having a direct manipulation interface between virtual slides in augmented reality (AR) space and users in the real world. This research has three subgoals; sharp slide images in lecture videos, direct manipulation on virtual slides, and hand-writable virtual slides using ordinary physical objects such as whiteboard markers. First, we implemented (a) an AR lecture recording system with an AR slide function, (b) a direct manipulation interface on AR slides, and (c) a handwriting function with whiteboard markers. Then, we conducted experiments in terms of system performance and slide visibility. As a result, the frames per second (fps) of our recording system is 30 fps or more. First, the result of the task-based user experiments achieved -17.85 seconds. Second, the questionnaire result of the seven-point likert scale improved +2.26 points. Finally, we concluded that our lecture recording system with a direct manipulation interface on AR slides is practical enough to improve the visibility of interactive slides. This paper describes the implementation and its evaluations.

Keywords: Augmented Reality, Lecture Video Making Support, Human Computer Interaction, Direct Manipulation, Image Processing

1 Introduction

Teaching styles are getting more diverse nowadays, therefore educational institutions require lecture support for online classes and flipped classrooms [1] [2]. Some lecturers provide lecture videos as lecture material distributed in advance. Some research [3] [4] show that high-quality lecture videos vary depending on the lectures. We consider the lecture video that displays clear and readable slides, lecturer's talking head, and interactive gestures with slides is high-quality.

Our purpose is to realize a system that can record the lecture video in which a lecturer interacts with clear slides, such as hand pointing and handwriting on slides. We consider

* Department of Computer Science, Graduate School of Engineering, Nagoya Institute of Technology, Aichi, Japan

some approaches to recording a lecture video of our objective. Some lecturers give a lecture in a lecture room with lecture slides displayed by a projector or a monitor. In such an environment, lecturers can easily perform body movements, such as hand gestures and non-verbal behaviors. However, viewers have difficulties in viewing lecture slides recorded in such physical-style lectures due to the camera angle, ambient light source, and a positional relationship of lecture contents in real space. We describe real lecture slide images caught on a camera as analog slides. In contrast, we describe slide images directly synthesized in a video as digital slides. In this paper, we make the difference between analog slides and digital slides. Digital slides provide stable and clear lecture content not influenced by the camera's surroundings. Another approach is to synthesize digital slides on the video with common presentation tools, such as PowerPoint and Keynote. With this method, lecturers have difficulties in direct manipulation on digital slides while giving a lecture in real space. We believe that lecture recordings that showcase the benefits of lectures with digital and analog slides are valuable lecture support. Our study explores novel and unobvious approaches that merge lecture recordings with digital and analog slides.

In this paper, we propose a lecture recording system equipped with a direct manipulation interface between digital slides in augmented reality (AR) space and lecturers in the real space. In previous research [5], we developed a system that uses AR technology and provides a digital lecture slide in AR space (hereinafter, called "AR slide"). The AR slide can display a clear lecture slide on the video. Lecturers have issues with the manipulation of digital slides. Hence, we aim to enable a lecturer to directly manipulate digital slides. Specifically, we implement a handwriting function that enables lecturers to handwrite on an AR slide with ordinary markers.

Our approach is a visualization of virtual objects in AR space (AR objects) with a projector. Our system creates lecture videos by combining AR slides and camera images using AR technology during the lecture. The lecturer must give a lecture while checking AR objects through created videos on the device screen, and proceeding with the two things simultaneously is very difficult. Because of that, we design a supportive environment for lecturers that projects the identical slide image onto the whiteboard in position, size, and content to the AR slide. In our system design, the lecturer visually gets the feedback of the AR slide and manipulates it through the projected slide image on an actual whiteboard. We implemented a writing function that enables lecturers to handwrite on an AR slide. Our system extracts handwritten content on the whiteboard and reflects it on the AR slide. Figures 1(a) and 1(b) show examples of ways to realize handwriting on an analog slide and a digital slide, respectively. We call these writing methods analog writing and digital writing. In this figure, analog writing is implemented by projected slides onto an actual whiteboard and writing on it directly. In addition, digital writing is implemented by capturing the screen of a writing tablet with digital slides. In contrast, we propose a writing method to enable lecturers to handwrite on a digital slide in the same way as analog writing. We call this method AR writing as shown in Figure 1(c).

In the experiments, we evaluate partial functions of our lecture recording system. Firstly, we showed the performance of our system. We checked if our system meets the minimum requirements for a lecture recording. Secondly, we conducted two experiments to evaluate the clarity and readability of handwritten slides. The first one is a recognition experiment of handwritten content on a slide using the task based on the Trail Making Test (TMT), which is a popular neuropsychological test [6] [7]. To the best of our knowledge, there are no conventional methods to evaluate the visibility of handwritten content by physical objects projected on virtual slides. Therefore, we designed the task in order to evaluate our

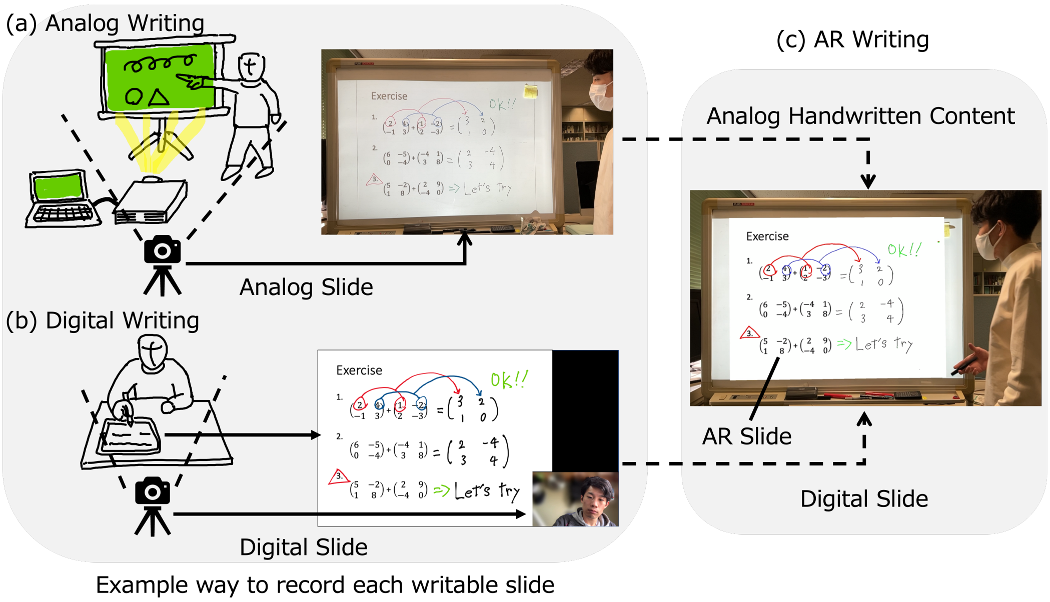


Figure 1: Analog, digital, and AR writing

handwriting function objectively and quantitatively. The second one is a subjective evaluation experiment by a seven point likert scale to evaluate the clarity and readability of handwritten slides. Based on these experimental results, we demonstrate the ease of viewing handwritten content and lecture slides by our system. The contributions of this paper can be summarized as follows:

- Realizing a lecture recording with virtual slides in a physical-style lecture environment using AR technology.
- Implementing a direct manipulation interface that enables lecturers to interact with virtual slides such as handwriting and hand pointing.
- A design of experiments based on TMT to show objectively and quantitatively that handwritten content on slides by our system is clear and readable.

The rest of the paper is as follows: Section 2 presents related works. Section 3 describes the AR lecture recording system, in particular, system design, AR slide, and the interaction with such a slide. Section 4 explains the implementation of our system. We explain our evaluation experiments with its results in Section 5 and discuss these results in Section 6. Finally, we conclude this paper in Section 7.

2 Related Work

We aim to realize a new supportive environment for recording lecture videos with AR and image processing technologies. Therefore, we investigated the effects of various teaching approaches on students' learning outcomes in educational institutions. Some researchers discuss the effects of interactive lectures by lecturers not only in face-to-face lectures [8] but also in lecture videos [3] [4]. From these studies, the high effectiveness of lecture videos is considered to vary depending on the subject, the learning level, and students' preferences. Among various types of lecture videos, the video of our objective shows the lecturer's talking head, clear lecture slides, and interactions between the lecturer and the slides.

Some studies use AR and 2D CG technologies and develop expressive and graphical user interactions in various situations such as educational support [9], video conference [10], and collaborative work [11]. User-friendly methods, such as a direct manipulation interface, for controlling visual effects based on gestures and postures of a presenter [12] or simple teaching motions [13] have been proposed. A user-friendly interface enables users make video effects intuitively without any programming or special skills.

We consider that interactive and graphical manipulations of presentation materials by a presenter lead to easier viewing of the materials and expressive storytelling for audiences. Saquib et al. [12] developed a direct manipulation interface that enables presenters to control digital graphs and figures in videos by body movements as input. Matulic et al. [14] proposed a projected avatar on presentation slides associated with the presenter's movements that interact with presentation elements and data. Our previous study [5] proposed a lecture recording with digital lecture slides displayed using AR technology. These applications realize digital presentation materials and interactive performance of the presenter, but they have issues with video feedback according to the manipulations. Due to these specifications, the presenters need to control the virtual materials while checking the video from a monitor. Therefore, this paper proposes recording of digital slides with a direct manipulation interface based on conventional and physical-style lectures in the real world.

The concept that realizes an experience of AR by projecting a virtual space state onto physical objects with a projector is known as spatial augmented reality (SAR). Some studies proposed approaches of visualization and interactions with virtual 3D objects based on SAR [15] [16]. Without SAR, AR users generally have to confirm virtual objects through their tablet device screen or head-mounted display. Third parties cannot experience that. Hartmann et al. [17] developed a wearable AR device that is equipped with a small projector to share the AR objects visually and work with another person collaboratively. Our approach has a similar concept with SAR which is based on projected images rather than tablet device or head-mounted display. Our system realizes a direct manipulation interface for the lecturer's interaction by projecting the same content, position, and size of the AR slide in AR space onto the whiteboard in real space. Particularly, we realize a writing function on an AR slide and a hand pointing.

In a writing function of our system, a lecturer handwrites on the projected whiteboard, and then handwritten content is reflected on the AR slide in AR space. We implement this function by image processing technology to extract the handwritten content. Our concept would also be implemented with special devices equipped with an image display and a painting function. For example, some researchers developed a digital whiteboard that enables remote users to write interactively on each other [18] and a writable LCD that supports collaborative work in multi-user scenarios [19]. However, our system uses an ordinary whiteboard and a projector to implement our writing function.

Table 1 summarizes the methods to realize handwriting functions on a slide. The methods vary depending on the types of slides and handwritten content. Based on the original media by which the handwritten content is extracted, we distinguish between digital handwritten content and analog handwritten content in this paper. Digital handwritten content is image data that is extracted from a device equipped with a writing function, such as an iPad or a digital whiteboard. In contrast, analog handwritten content is written on a physical

Table 1: Analog/digital slide and handwritten content

Slide \ Handwritten content	Digital handwritten content	Analog handwritten content
Digital slide	Handwriting with iPad and PowerPoint. (Digital writing)	Handwriting on an ordinary whiteboard and reflecting it on a digital slide (AR writing)
Analog slide	Handwriting on a digital whiteboard and projecting an analog slide	Handwriting on an ordinary whiteboard and projecting an analog slide (Analog writing)

object, such as a whiteboard or a blackboard. In this paper, the method to realize the digital handwritten content on a digital slide is called digital writing, whereas the analog handwritten content on an analog slide is called analog writing. We realize analog handwritten content on a digital slide in AR space, which is called AR writing. A method to realize digital handwritten content on an analog slide is not considered in this paper because it requires a writable device and additional efforts than digital writing.

3 AR Lecture Recording System

3.1 System Design

In this paper, we aim to develop a lecture recording system with a direct manipulation interface for interactions between virtual slides in AR space and lecturers. The previous research [5] proposed the presentation system with the AR slide, however, the presenter has difficulties in interacting with virtual content without any feedback and support. Hence, we improve the AR slide function and design a new system to achieve the following three subgoals of our study:

- clear lecture slides in lecture videos,
- direct manipulation of virtual slides by lecturers, and
- hand-writable virtual slides with ordinary whiteboard markers.

Figure 2 shows our system, called AR lecture recording system. This system consists of an AR recording device, a remote operation device, a lecture Personal Computer (PC), and a projector. AR recording device plays a role of a camera and records lecture videos. We assume that the lecturer gives a lecture by him/herself with an AR recording device placed at a fixed position. Using AR technology, the AR recording device combines real camera images with virtual objects in AR space, such as an AR slide and a digital pointer. It is difficult for a lecturer to directly operate an AR recording device or a lecture PC. Because of this, the lecturer remotely operates our system during lecture recording via a remote operation device.

A lecture PC displays projected images to show the state of digital objects in AR space, and a projector projects the image onto a whiteboard in real space. The lecturer cannot

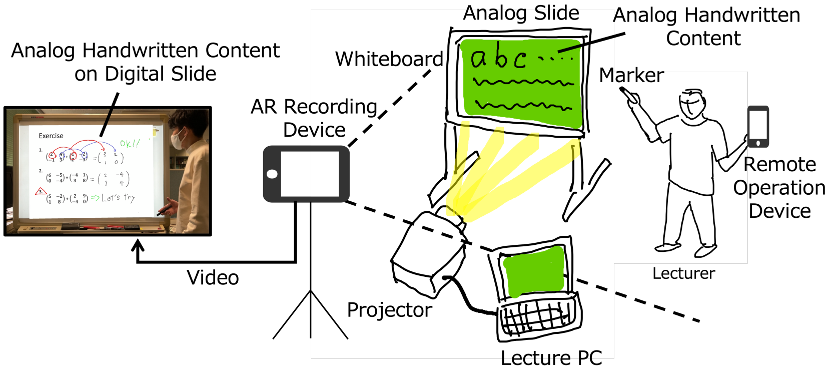


Figure 2: System design

directly see an AR slide, which is an AR object only displayed on the device, but can only see them through the device screen. In our previous system [5], the lecturer could confirm lecture slides from another prepared display monitor that mirrors a screen of the AR recording device. However, the lecturer had to simultaneously proceed with a lecture and check the mirrored monitor; for example, confirming the content of lecture slides and pointing to the slide. The system could be a burden for a lecturer. Therefore in this work, we made our system to enable the lecturer check the content of AR objects and interact with the AR slides through the actual projected images.

3.2 AR Slide

In our system, the lecturer gives a lecture using AR slides. AR slide is a virtual object placed three-dimensionally in AR space that displays a lecture slide. AR slide is directly rendered in the real camera image without a display monitor, a projector, and video editing after recording. Thus, an AR slide can display a clear and stable slide not influenced by the camera angle and ambient light source.

Our system provides the function to place an AR slide at any position using AR technology. The plane detection of AR technology can recognize the shape of the real space such as vertical planes. Our system can place the AR slide as if it were pasted on the vertical plane, which is recognized by the plane detection. Consequently, the lecturer is able to automatically place the object at a desired position in the real three-dimensional space without manually setting the position and orientation. In addition, an AR slide is a virtual object, thus the size and aspect ratio can be freely changed. Therefore, our system enables the lecturer to place the AR slide in the same position and size as superposed on the projected slide image on a whiteboard in real space.

Figure 3 shows the positional relationship of an AR slide, a projected slide on a whiteboard, and a lecturer. This represents a positional relationship not in real space but in the video recognized by our system. In the video, a lecturer, an AR slide, a projected slide, and a whiteboard are placed in order of proximity to the camera. AR slides do not exist in the real space, hence the system has to render these positional relationships as video images. Our system uses the people occlusion to render digital objects in accordance with these positional relationships. The people occlusion is an AR function that recognizes a front-back relationship between people and AR objects appropriately depending on the depth map.

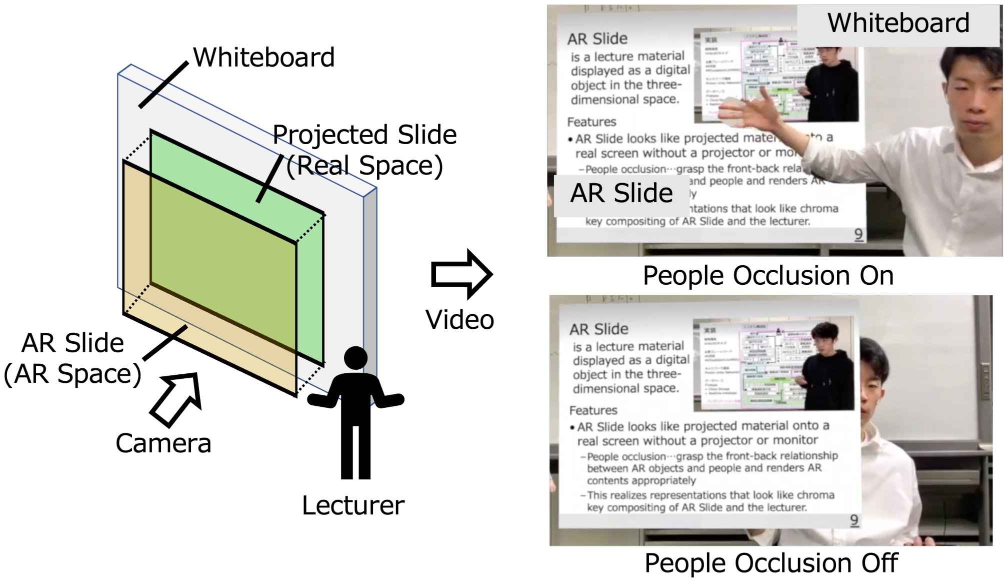


Figure 3: Left: System environment, Right: People occlusion

Therefore, the spatial relationship between AR objects and the physical objects is rendered in the video. With the people occlusion, the lecturer is not hidden, whereas the projected image and written content on a whiteboard are hidden by the AR slide shown at upper right in Figure 3. As an example of this mechanism usage, a projector is able to project the lecture manuscript without appearing it in lecture videos.

3.3 Interaction with AR Slides

Our system mainly provides two interactions: handwriting on an AR slide and hand pointing. The handwriting function of our system enables lecturers to handwrite on an AR slide, which is a virtual object in AR space, without any writing tablets. The same lecture slide image as the AR slide is projected on a whiteboard. Hence, the lecturer visually confirms the slide content and writes on a whiteboard. Figure 4 denotes the handwriting function which reflects the written content on a whiteboard to the AR slide as digital data. The lecturer can therefore write shapes and letters on the AR slide with markers as if he/she were writing on a whiteboard. Our system supports four colors of whiteboard markers: black, red, blue, and green.

The lecture recording used by our system provides two ways of pointing to an AR slide as shown in Figure 5. The first way is the digital pointer that indicates a position by displaying a red circle or arrow on the AR slide. In the previous study, we developed a digital pointer function using a remote operation device [5]. With the remote operation device displaying a lecture slide image as an interface, our system displays a digital pointer at the touched position. However, this pointer function is not sufficient for interaction with AR slides. The lecturer had difficulties in operating the digital pointer due to the necessity of checking the device screen or getting the feedback of an operation. Hence in our system, a lecture PC displays a pointer at the same position in the projected slide depending on the operation. The lecturer is able to operate a digital pointer while getting the feedback in real

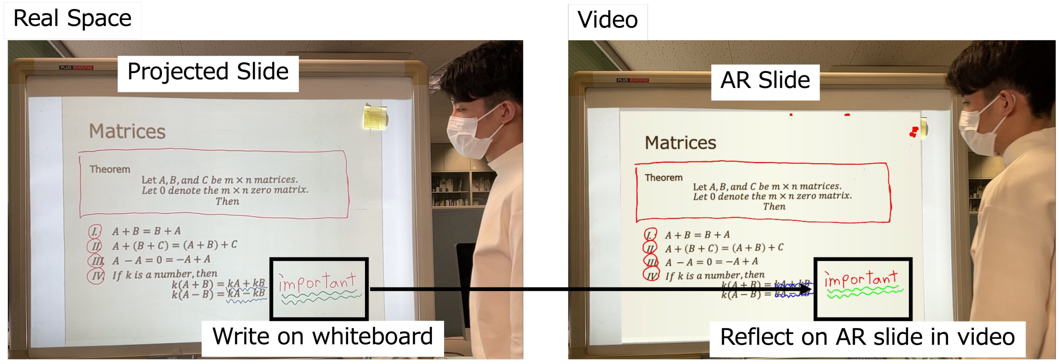
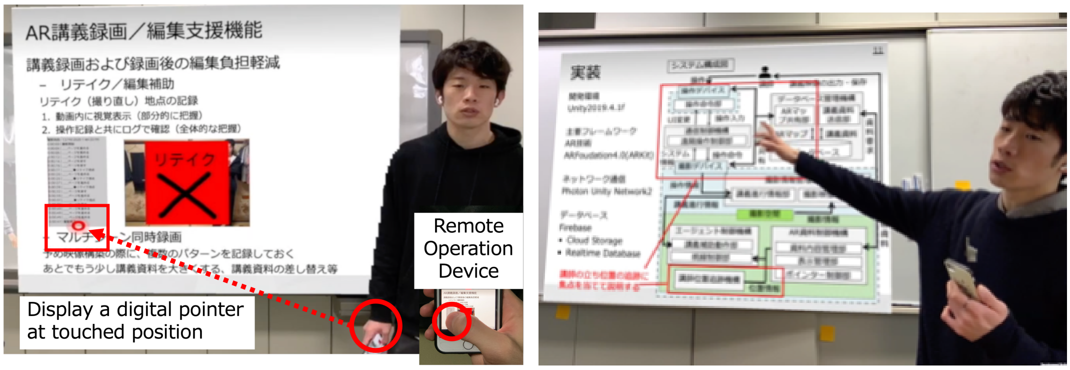


Figure 4: Handwriting on a virtual and digital slide (AR slide)



(a) Digital pointer (b) Hand pointing

Figure 5: Two ways of pointing to an AR slide that our system provides

space.

The second is a hand pointing. We assume that the lecturer performs a hand pointing depending on the lecture's progress. For example, it is when explaining lecture slides with body gestures. In addition, it is cumbersome for lecturers to operate a digital pointer with a device while writing on lecture slides. The people occlusion enables a system to express such as chroma key compositing of an AR slide and a lecturer, thereby realizing a hand pointer. Our system projects the slide image onto a whiteboard that is synchronized with the position, size, and content of the AR slide. Thus, the lecturer can perform a hand pointing to the AR slide depending on the projected slide.

4 Implementation

In this section, we describe the implementation of our system architecture and functions. Ito et al. [5] explained the details on the implementation of the AR slide. An AR recording device and a remote operation device were developed as iOS applications. The development framework is Unity 2019.4.1f and the development language is C#. A lecture PC was developed as Web application. The development language is JavaScript with React 17.0.2, which is a user interface library.

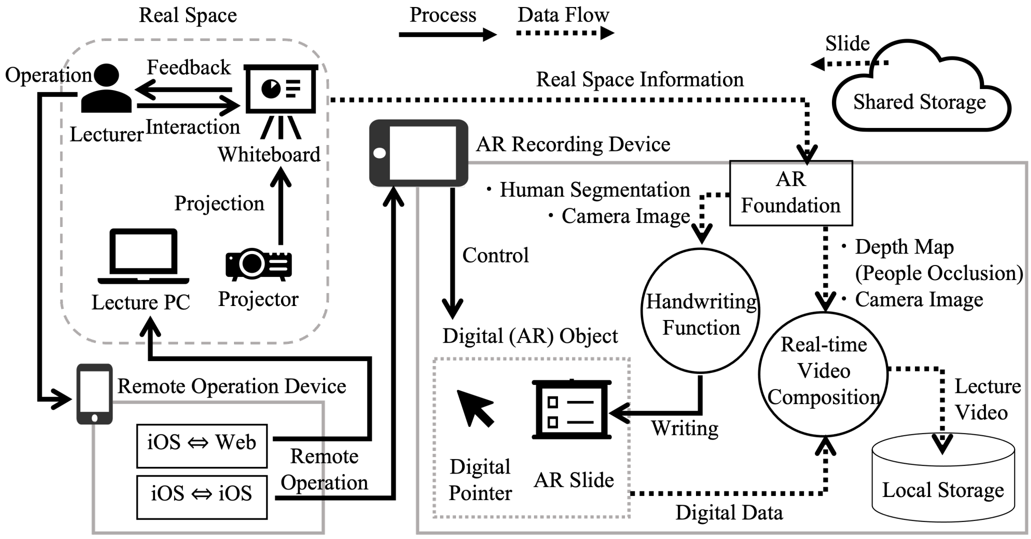


Figure 6: System architecture

4.1 System Architecture

Figure 6 shows our system architecture. A lecture PC controls the projected image by the projector as Web application. An image is projected on a whiteboard. The lecturer receives the feedback on his/her system operation based on the projected image. A lecture PC mainly controls two projected contents: lecture slides and a digital pointer. The projected slide is synchronized with the AR slide content. This device also adjusts the brightness and opacity of the projected slide depending on the handwriting function. When the lecturer uses a digital pointer on an AR slide with the system function, a pointer image is projected at the same position in the projected slide. In this way, the lecturer proceeds with lecture recording receiving the feedback of the AR slide content and the digital pointer.

Lecturers manipulate pages of lecture slides and a pointer to proceed with lecture recording. In our system, a lecturer has difficulties in directly operating the AR recording device placed several meters away. Therefore, the lecturer can operate the system remotely using a remote operation device in hand. The remote operation device controls a lecture PC and an AR recording device simultaneously depending on the operation by a lecturer. The remote operation device has two sub-modules in accordance with the communication group of the devices. We used the communication framework Photon Unity Networking 2[1] for the communication function between the iOS applications, although Realtime Database by Firebase[2] is used for between the Web application and the iOS application. The shared storage contains the lecture slides used in each device.

AR recording device constructs the lecture video and records it to the local storage in this device. Our system is a lecture recording system using AR and image processing technologies. We implemented AR functions using AR Foundation 4.1.7[3], which is an AR development framework for iOS in Unity. AR lecture device uses AR technology to acquire various kinds of information in the real space and construct videos on the basis of this infor-

[1] https://www.photonengine.com/pun
[2] https://firebase.google.com
[3] https://unity.com/unity/features/arfoundation

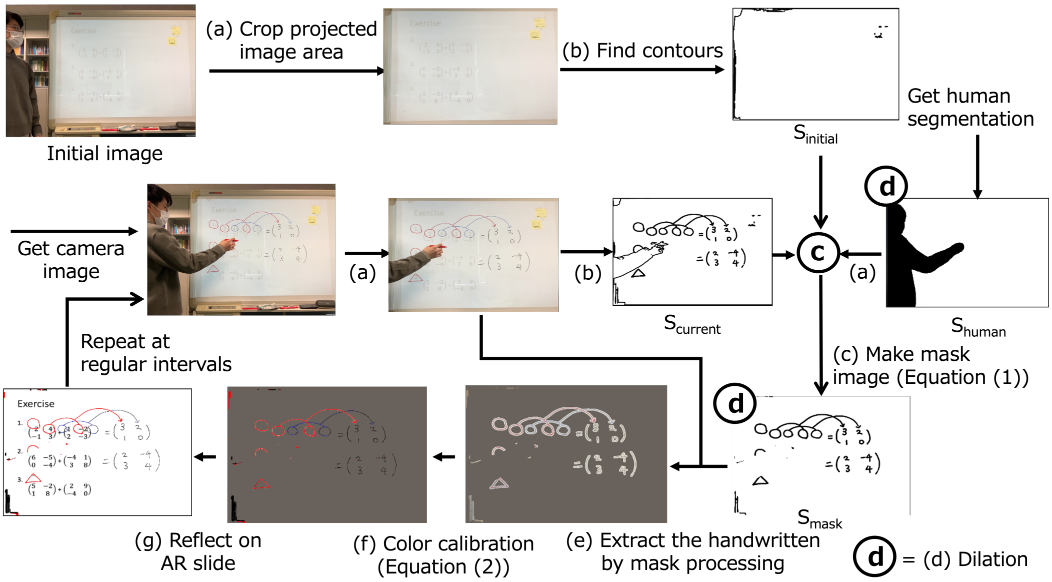


Figure 7: Handwriting mechanism

mation. The real space information includes a real camera image, the human segmentation image to identify the area of the lecturer in a camera image, and a depth map to measure the distance between the lecturer and the AR recording device. The AR recording device controls digital objects in the AR space, such as a digital pointer and a handwritten AR slide. The people occlusion can grasp and render the front-back relationship between people and AR objects appropriately. A depth map is an image channel that contains information relating to the distance from the device to physical objects. AR Foundation estimates the distance between a human and a device using this depth map. Finally, the AR recording device creates lecture videos during the lecture by combining real camera images and AR objects at the appropriate position.

4.2 Handwriting Function

This section describes the implementation of a handwriting function as direct manipulation of AR slides. This function enables the lecturer to handwrite on the AR slide in AR space by writing on the whiteboard in real space. Our system processes images with OpenCV 3 [4], which is an image processing software library.

Figure 7 shows the process flow of this function. Process (a) crops only the projected slide area on an actual whiteboard from a camera image acquired with an AR recording device. The projected slide is identical with the AR slide in terms of position, size, and content. Since an AR slide is object data in the system, it provides its own three-dimensional (3D) coordinates. Therefore, the system is able to calculate the area of the projected slide in a camera image based on the 3D coordinates of the four AR slide corners. At first, the 3D coordinate of each AR slide corner is changed into the coordinate of a camera coordinate system with respect to a camera position as the origin. A scene view is rendered by projecting 3D coordinates into the image plane using a perspective transformation. The

[4]https://opencv.org

view frustum of this system represents the region of camera rendering, and its shape is a trapezium. Thus, the system needs to convert a coordinate in the view frustum into the normalized device coordinate by a perspective transformation and a normalization. Finally, the system calculates the coordinates of the AR slide in a camera image multiplying the normalized value by the acquired image size.

Process (b) detects the contour lines of the images in order to identify the handwritten content. Our system uses the border tracing algorithm as Suzuki85's algorithm [20] to extract the contours of the handwritten content on a whiteboard. The slide image is projected on the whiteboard in our system, thus this method may also detect not only the handwritten content but also the projected image. Hence, a lecture PC makes the brightness and opacity of the projected image lighter than usual. In a preliminary experiment, we verified that an error detection was sufficiently reduced enough for the lecturer to recognize the projected contents. We also expect that the contours extraction with the projected image as a background leads to the extraction not affected by a surrounding environment, such as ambient light source and a whiteboard type.

It is assumed that our system even detects the contours not related to the handwritten content, such as scratches and frames of a whiteboard. In addition, we need to prevent the system from extracting the lecturer, who is writing on the whiteboard, as contours. Therefore, the process (c) removes such irrelevant contours to the handwritten content and generates a mask image. Our system stores an initial image in advance before the lecturer begins to write. The initial image is acquired when the handwriting function is enabled. When creating a mask image, the system extracts only the handwritten content by detecting the contours difference between the initial image and acquired image at every process flow. This process is implemented owing to the specification of our system that the camera placed at a fixed position. A human segmentation image is a binary image that segments the pixels in the image into human regions and the other regions. Using the human segmentation image, the system also removes the regions of the lecturer from the mask image. $S_{initial}$, $S_{current}$, S_{human}, and S_{mask} are all binary images. $S_{initial}$ and $S_{current}$ respectively denote contour features of the initial image and acquired image at the current process. S_{human} is the human segmentation image. S_{human} and S_{mask} are dilated by process (d). Process (c) is formulated as the following Equation (1).

$$S_{mask} = S_{current} \wedge \neg(S_{initial} \vee S_{human}). \tag{1}$$

Process (e) extracts the handwritten content from the camera image by the mask image S_{mask} created on the process (c). The extracted image is an RGBA image with opacity parameter added to an RGB image, which is transparent except for the handwritten content. This image just reproduces colors of the written content on a whiteboard in a camera image. We anticipated that the contrast between the handwritten content and the clear AR slide would affect the ease of viewing. For this reason, the process (f) is a color calibration phase that identifies and converts the colors, and thus unifies the colors (RGB parameters) according to each color. The color calibration processes the conversion of the color for each pixel based on the RGBA function (Equation (2)).

$$RGBA_{dst(x,y)} = \begin{cases} (*,*,*,0) & \text{if } InRange(HSV_{src(x,y)}, (0,0,180), (180,30,255)) \\ (255,0,0,255) & \text{else if } InRange(HSV_{src(x,y)}, (0,50,80), (30,255,255)) \parallel \\ & InRange(HSV_{src(x,y)}, (150,50,80), (180,255,255)) \\ (0,255,0,255) & \text{else if } InRange(HSV_{src(x,y)}, (30,30,50), (90,255,255)) \\ (0,0,255,255) & \text{else if } InRange(HSV_{src(x,y)}, (90,30,50), (150,255,255)) \\ (0,0,0,255) & \text{else if } InRange(HSV_{src(x,y)}, (0,0,0), (180,255,180)) \\ RGBA_{src(x,y)} & \text{else} \end{cases} \tag{2}$$

We define the *InRange* function provided that each argument is a three dimensional array and *lb* and *ub* are lower and upper bounds, respectively.

$$\begin{aligned} &InRange((H,S,V),(lb_0,lb_1,lb_2),(ub_0,ub_1,ub_2))) \\ &= (lb_0 \le H \le ub_0) \wedge (lb_1 \le S \le ub_1) \wedge (lb_2 \le V \le ub_2). \end{aligned} \tag{3}$$

Our system identifies four whiteboard marker colors (red, blue, green, and black) and a background color (white) based on the HSV parameters of each pixel. According to the OpenCV specification, the hue of HSV space takes values in [0,180], whereas the others are in [0,255]. We tuned these parameters empirically. White color is converted to transparent (alpha value = 0), and other colors are converted to unified color. If the system fails to identify a color, it does not convert the color.

Process (g) uses the calculated coordinates of the AR slide on a device screen and displays the handwritten content onto the AR slide. The handwriting function of the system repeats these processes at regular intervals. This system updates 30 times per second.

5 Evaluation

In this section, we describe the experiments to evaluate partial functions such as system performance and handwritten slide visibility. Our system performs video processing during the lecture to create a lecture video that shows the lecturer's interactions with digital lecture slides in virtual space. The evaluation experiment consists of the following two phases. At first, we verified that the performance of our system meets the minimum requirements for a lecture recording. And then, we designed the test and evaluated handwritten content and slides by our system in terms of clarity and readability.

The execution environment of this paper is as follows: iPad Pro 3rd generation running on iOS 14.8.1 as AR recording device and iPhone 8 running on iOS 14.6 as remote operation device. We used a MacBook Pro running on macOS Big Sur 11.6 as lecture PC and RICOH WX4152, which is a short focus projector.

5.1 System Performance

The feature of our system is to use AR and image processing technologies for real-time video processing without any video editing after recording. We defined minimum requirements for the system as 1) the lecturer records the lecture by him/herself, 2) the system requires no special equipment, such as a digital whiteboard and 3) stable and real-time video processing. In Section 3.1, we described our system design considering 1) and 2). Specifically, the requirements 3) are that the system can record videos at 30 fps or more

Table 2: Rate of rendering in frames per second

	before (10s)		after (10s)	
	Min [fps]	Mean ± SD [fps]	Min [fps]	Mean ± SD [fps]
Slide page operation	59.80	59.98 ± 0.04	59.77	59.98 ± 0.04
Digital pointer	59.69	59.98 ± 0.05	57.98	59.97 ± 0.12
Handwriting function				
・10 times per second	59.83	59.98 ± 0.03	48.99	51.22 ± 0.97
・30 times per second	59.90	59.98 ± 0.01	35.50	36.02 ± 0.13
・60 times per second	59.88	59.98 ± 0.02	26.20	30.16 ± 1.66

and provide low latency for the communication between each device. In this section, we verified 3).

Video processing requires less processing load. If the processing load is heavy, the lecture video might be lagging or the system might cause errors. Fps stands for frames per second and implies the degree of processing load. Hence, we measured the minimum, average, and standard deviation of the fps values for 10 seconds before and after the use of each function, respectively. We measured the handwriting function in three different cases in accordance with the maximum times of update per second: 10 times per second, 30 times per second, and 60 times per second. We assume one process flow of the handwriting function in Figure 7 is one time. The number of trials for each function is 30. The measurement fps results are shown in Table 2.

The handwriting function recorded an average of 30.16 fps, which is the highest processing load among the functions. Video contents such as TV, movies, and DVDs often use 30 fps or less videos. Therefore, the results sufficiently indicated the stability of the video processing of our system. Maintaining a stable fps value leads to easy watching of the video.

The communication among the devices of our system requires low latency in order to realize the interaction between the lecturer and digital lecture slides. The real-time performance of the communication is required both between two iOS applications and between Web and iOS applications. A delay in the processing of an AR recording device from a remote operation device may cause a discrepancy between the images as intended by the lecturer and the actual constructed images. A delay of lecture PC leads to a delay of the feedback of a lecturer's manipulation on the lecture progress, hence may affect the usability of the system.

We measured the response time of system functions in order to demonstrate the real-time performance of our system communication. The response time is the elapsed time from the operation command by a remote operation device until the target device starts its processing. We measured 100 times for each function. The average ping of the runtime environment was 37.0 ms. Table 3 shows the response time results. The results denote that our system has enough real-time communication for lecture recording without any problems. The response time of a projector, particularly the page operation and handwriting functions, was lower than the one of an AR recording device. Due to the specifications of our system, the lecturer basically conducts the lecture based on the projected lecture slides. Eventually, the lecturer is able to clip the unnecessary video parts caused by the delay in the progress of lecture. Therefore, we considered that a slight delay of the projector would

Table 3: Response time

	Min [ms]	Max [ms]	Mean ± SD [ms]
Remote operation ⇒ AR recording device			
Slide page operation	33	462	129 ± 86
Digital pointer	31	295	126 ± 58
Handwriting function	26	337	147 ± 61
Remote operation ⇒ Projector			
Slide page operation	119	603	406 ± 83
Digital pointer	148	262	200 ± 35
Handwriting function	141	784	332 ± 115

not pose any significant practical problem. From these results, we suggest that our system has enough performance to meet the minimum requirements.

5.2 Video Production

5.2.1 Procedure

We conducted two experiments in this study. The first experiment is a recognition experiment to quantitatively evaluate the degree of identification of handwritten content in a practical task. The second experiment is a quantitative experiment for subjective evaluation of the clarity and readability of handwritten slides. In our experiments, participants first perform a recognition experiment, and then answer a questionnaire for subjective evaluation. We compared the handwritten slides created by the following four methods.

1. Digital writing: Using PowerPoint's drawing tool, digital handwritten content on a digital slide is converted into an image and then reconstructed in AR space. Although this method is not possible in the actual lecture recording, we used this as one of the comparison methods.

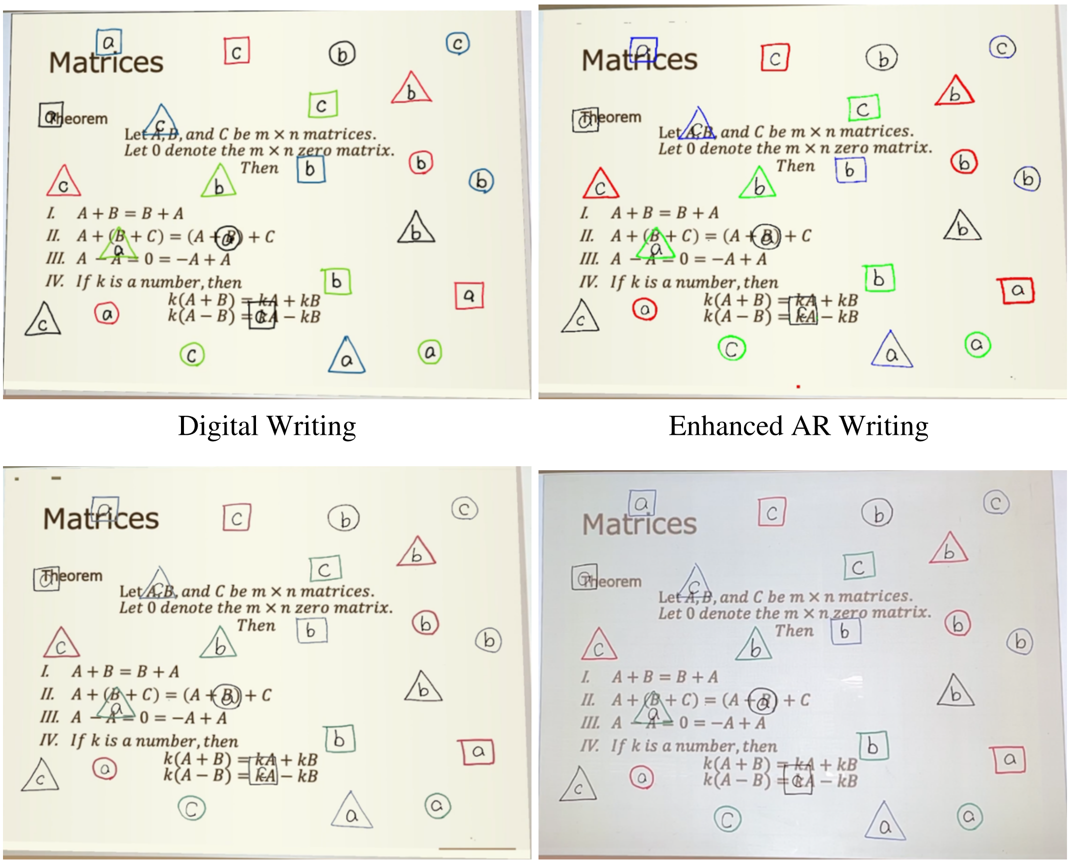
2. Enhanced AR writing: Handwriting function of our system (Figure 7).

3. AR writing: An analog handwritten content on the whiteboard is extracted and reproduced on an AR slide. This method is our handwriting function without a color calibration phase.

4. Analog writing: A method of shooting analog handwritten content on an analog slide on a whiteboard.

We assume that the differences of test data by each method in the writing style on preparing it would affect the results. Hence, we first made the test data by digital writing, and then traced over it to make test data by the others.

Here is a description of the recognition experiment we made. In our recognition experiment, we use the Trail Making Test (TMT) to evaluate handwritten lecture slides. TMT is one of the most popular neuropsychological tests that measures visual cognitive functions, such as visual tracking and processing [6]. TMT can be applied to measure mental flexibility, attention functions, and the ability to recognize numbers and letters [7]. TMT is a test in which participants trace the symbols on a sheet in a predetermined order. These



Digital Writing | Enhanced AR Writing

AR Writing | Analog Writing

Figure 8: Examples of recognition test sheets for four methods

symbols consist of numbers and letters generally. The degree of visual cognitive function is evaluated based on the completion time of TMT.

In our recognition experiments, participants conduct the same task as TMT. The identified symbols on the test sheet are handwritten letters and symbols by each comparison method as shown in Figure 8. The test sheets are cutouts of the target slide from the movie. We assume that these methods are evaluated as writing methods on the situation of lectures. Hence, a lecture slide is overlaid with symbols on the test sheet as the background image. We prepared five types of lecture slides for the background slide.

Our test sheets contain 24 symbols randomly selected from 36 symbols with three letters (a, b, and c), four colors (red, blue, green, and black), and three shapes (square, triangle, and circle). The participants should quickly trace the 24 symbols in the predetermined order we gave. We estimate the recognized degree of the handwritten content on lecture slides from the completion time of the TMT.

Participants perform four recognition tests with different test sheets prepared by each method. In all four tests, the arrangement of the symbols on the test sheets is different. We created 80 test sheets using four comparison methods, four different orders, and five different background slides. The background slide is randomly selected from the five slides before the test, and then it is used throughout all the four tests. Similarly, the method and the order of the four tests are also randomly selected. Since participants perform the recognition test four times consecutively, the habituation to the test may affect the experimental results depending on the order of methods. Therefore, we took the randomness and fairness of the experiment into account and randomized the order. After the recognition test, we collected feedback and opinions from the participants. We told them before the test that we would give prizes to the top performers.

Next, we describe the subjective evaluation experiment. In the subjective evaluation, participants anonymously answer a questionnaire regarding the ease of viewing lecture slides handwritten by four different methods. The participants are not informed of the details of the methods and which method is selected throughout the experiment. In this experiment, we used the same five slides that are used in our recognition experiments as lecture slides. Moreover, we made the handwritten content according to the content of the lecture slides. We prepared 20 handwritten slides with five types of lecture slides and the four methods. The participants evaluate each handwritten lecture slide by a seven point likert scale (from “difficult viewing” to “easy viewing”).

5.2.2 *Results*

This section shows the results of the two evaluation experiments. In these experiments, 8 undergraduate and 6 graduate students participated. Our system is a lecture recording system intended for educational institutions, therefore only students performed the experiments. In consideration of the experimental design of our recognition test, the target participants were students with no cognitive problems.

Table 4 denotes the results of the recognition test. The shorter completion time means that the handwritten symbols are more easily identified in practical tests such as recognition tasks. Comparing the average completion times, the best result was digital writing, enhanced AR writing (ours), AR writing, and analog writing, in the order. We conducted a t-test to compare the digital writing with the others and found significant differences in task completion time for both AR writing and analog writing. Nevertheless, the difference between digital writing and enhanced AR writing was not statistically significant. The estimated 95% confidence interval for the difference in completion time between these methods was $[-11.92s, +3.49s]$. These results indicate that our method is as effective as digital writing in practical cognitive tasks.

Figure 9 represents the results of the subjective evaluation experiment with a questionnaire of a seven point likert scale. The higher score means that the students evaluated it better as a handwritten slide regarding ease of viewing. In the figure, the horizontal axis contains the category of each slide and total, and the vertical axis shows the average of its scores. Comparing the seven point likert scores, digital writing, enhanced AR writing, analog writing, and AR writing were observed in descending order of the score. The results of each slide had a tendency to be roughly the same as the total score. The results indicated that our method had the same high subjective evaluation as digital writing.

As the results of both objective and subjective evaluations, we found that our method performed as well as the one used in PowerPoint. Moreover, our method showed superior performance in terms of clarity and readability to AR writing and analog writing.

6 Discussion

In this section, we discuss the usability of our lecture recording system using AR technology. Our study proposes a lecture recording that includes AR slides, which are digital slides in AR space, and interaction with them. We implemented the handwriting function

Table 4: Experimental results (Trail Making Test)

	Min - Max [s] - [s]	Mean ± SD [s]	t-test
Digital Writing	39.34 - 78.31	61.04 ± 11.33	-
Enhanced AR Writing	38.62 - 100.06	65.26 ± 18.38	$p = 0.129$
AR Writing	63.10 - 156.47	112.24 ± 30.81	$p = 2.489 \times 10^{-6}*$
Analog Writing	43.53 - 147.47	84.06 ± 27.87	$p = 2.421 \times 10^{-4}*$

* significant difference found by t-test, one sided test, $p < 0.05$

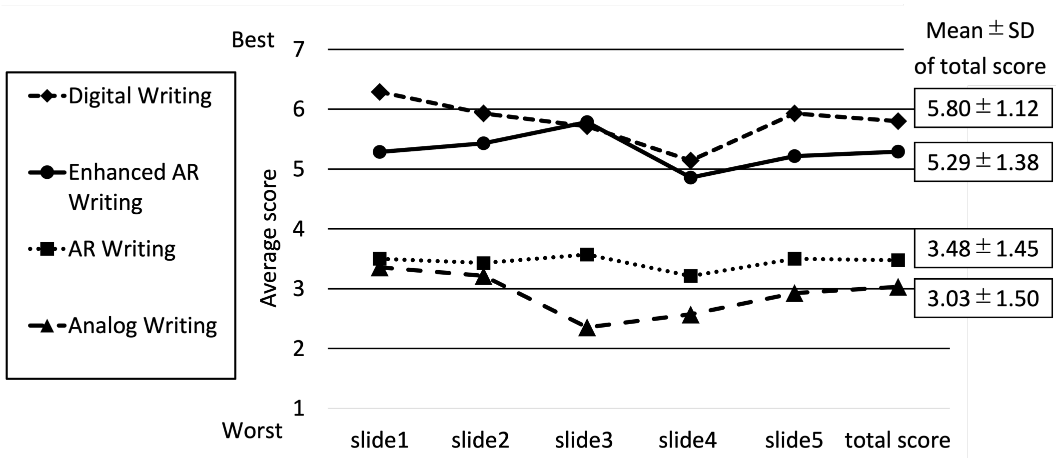


Figure 9: Experimental results (Questionnaire)

on an AR slide in AR space as direct manipulation. Table 4 denotes enhanced AR writing (our method) has no significant difference in its performance compared to digital writing in practical tasks such as recognition experiments. The difference of averages in the recognition experiment between enhanced AR writing and digital writing was about 4 seconds. From the results of the subjective evaluation by a seven point likert scale, enhanced AR writing was slightly 0.51 points lower than digital writing as shown in Figure 9. Thus, the results of both objective and subjective evaluations demonstrate that our handwritten function performs almost equally compared to digital writing. Digital writing just displays the images created in advance with PowerPoint. In order to realize this method during the lecture, the lecturer requires another device, such as a writing tablet. However, it is a burden for lecturers to conduct the lecture and manipulation in such a method. On the other hand, our system provides a handwriting function on the AR slide by simply writing on the whiteboard. Besides, the performance of our method obviously showed a superior result to other methods: AR writing and analog writing. Therefore, our system enables the lecturer to write by hand in a practical method while providing clear and readable slides. In addition, we verified that the function of adjusting the position and size of an AR slide in order to overlay on a projected image performs sufficiently in preparing the handwritten slide images for the experiments. Since a camera is fixed, the lecturer only needs to set an AR slide on a projected image so as not to be misalignment once before lecture recording. Therefore, the proposed method showed a similar quality to digital slides in a physical-style presentation such as analog writing.

Our handwriting function first extracts the handwritten content and then calibrates its colors. In our experiment, we compared this handwriting function without color calibration as AR writing. The result of AR writing showed a low score, whereas enhanced AR writing is high. From these results, we consider that the color calibration process influences the ease of viewing the video content. In addition, some participants remarked that AR writing had difficulties in distinguishing colors, in particular blue, black, and green. They did not mention the identification of shapes and letters. These results indicate that it is not enough to simply reproduce the extracted handwritten content on the digital slide if color markers are used.

The difference between AR writing and analog writing was observed to some extent, even though the images use the same handwritten content according to the implementation. This is because we consider the contrast between the brightness of an AR slide and the handwritten content prevents the identification of colors. Although color calibration is an important process in our system that provides the AR slide as lecture content, it can cause errors in identifying the colors of handwritten content. Our system supports red, blue, green, and black as marker colors. If there are any problems with this color calibration, we should examine another method of identifying the colors or reducing supported colors.

We implemented the extraction of handwritten content empirically, such as detecting the contour lines and parameter tuning for color calibration. We will also examine another method based on machine learning [21] in the future, although it must be a lightweight learning model for mobile devices. In this paper, we implemented our handwriting function in consideration of the processing load for real-time video processing.

Our experimental results demonstrate that AR slides improve the visibility of digital video content for lectures. Enhanced AR writing and digital writing performed significantly better than the others. These methods digitize the lecture slides. Comparing AR writing and analog writing, AR writing showed a better score (average score +0.45 points) in the subjective evaluation, although analog writing had a better result (average time -28.18s) in the recognition test of handwritten content. We suppose that the evaluation of the AR slides themselves affects the subjective evaluation positively. Therefore, our concept of digitalizing video content is effective enough to create a lecture video with clear and readable slides.

7 Conclusion

We developed a lecture recording system that enables readable and interactable slides in AR spaces on general-purpose devices. In other words, our system provides a direct manipulation interface between digital slides in the AR space and lecturers in the real space. In particular, our system provides a handwriting function and a hand pointing as the lecturer's interaction. Specifically, our handwriting function enables users to write on AR slides using actual whiteboard markers directly because we use a whiteboard to display slides by a projector and write on the AR slides. Furthermore, this paper showed how to extract handwritten content from the whiteboard displaying a slide. Then we conducted experimental evaluations on the system performance and the visibility of slides. The performance test results show that our system is adequate for making 30 or more fps videos using AR technology. In addition, the user test results show objectively and quantitatively that the digital slides handwritten by our system can display clear and readable lecture slides in the video.

As a result, our system can help lecturers make lecture videos that include interactions between lecturers and clear slides. This system contributes to spreading high-quality lecture videos for a wide variety of online learning environments.

Acknowledgments

This work was supported in part by JSPS KAKENHI Grant Numbers JP19K12097, JP19K12266, and JP22K18006.

References

[1] M. Furini, G. Galli, and M. C. Martini. An online education system to produce and distribute video lectures. *Mobile Networks and Applications*, 25:969–976, 2020.

[2] H. Okumoto, M. Yoshida, K. Umemura, and Y. Ichikawa. Response collector: A video learning system for flipped classrooms. In *5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA)*, pages 176–181, 2018.

[3] P. J. Guo, J. Kim, and R. Rubin. How video production affects student engagement: An empirical study of mooc videos. In *Proceedings of the First ACM Conference on Learning at Scale Conference*, page 41–50. Association for Computing Machinery, 2014.

[4] Y. Y. Ng and A. Przybyłek. Instructor presence in video lectures: Preliminary findings from an online experiment. *IEEE Access*, 9:36485–36499, 2021.

[5] Y. Ito, M. Kikuchi, T. Ozono, and T. Shintani. Developing a lecture video recording system using augmented reality. In *10th International Congress on Advanced Applied Informatics (IIAI-AAI)*, pages 65–70, 2021.

[6] T. N. Tombaugh. Trail making test a and b: Normative data stratified by age and education. *Archives of Clinical Neuropsychology*, 19(2):203–214, 2004.

[7] S. F. Crowe. The differential contribution of mental tracking, cognitive flexibility, visual search, and motor speed to performance on parts a and b of the trail making test. *Journal of Clinical Psychology*, 54(5):585–591, 1998.

[8] Y. Shi, H. Yang, J. Zhang, S. Wang, and H. H. Yang. The effects of interactive whiteboard-based classroom instruction on students' cognitive learning outcomes: A meta-analysis. In *International Symposium on Educational Technology (ISET)*, pages 276–280, 2019.

[9] S. Pochtoviuk, T. Vakaliuk, and A. Pikilnyak. Possibilities of application of augmented reality in different branches of education. *Electronic*, pages 179–197, 2020.

[10] I. Barakonyi, T. Fahmy, and D. Schmalstieg. Remote collaboration using augmented reality videoconferencing. In *Proceedings of Graphics Interface 2004*, page 89–96, 2004.

[11] N. H. Lehment, K. Erhardt, and G. Rigoll. Interface design for an inexpensive hands-free collaborative videoconferencing system. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 295–296, 2012.

[12] N. Saquib, R. H. Kazi, L.-Y. Wei, and W. Li. Interactive body-driven graphics for augmented video performance. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, page 1–12.

[13] H. Kataoka, T. Ozono, and S. Toramatsu. Developing an AR pop-up picture book and its effect editor based on teaching motions. *International Institute of Applied Informatics*, 7(1):1–10, 2021.

[14] F. Matulic, L. Engeln, C. Träger, and R. Dachselt. Embodied interactions for novel immersive presentational experiences. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, page 1713–1720.

[15] H. Benko, R. Jota, and A. Wilson. Miragetable: Freehand interaction on a projected augmented reality tabletop. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, page 199–208, 2012.

[16] A. D. Wilson and H. Benko. Combining multiple depth cameras and projectors for interactions on, above, and between surfaces. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology*, page 273–282, 2010.

[17] J. Hartmann, Y.-T. Yeh, and D. Vogel. AAR: Augmenting a wearable augmented reality display with an actuated head-mounted projector. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, page 445–458, 2020.

[18] K. Higuchi, Y. Chen, P. A. Chou, Z. Zhang, and Z. Liu. Immerseboard: Immersive telepresence experience using a digital whiteboard. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, page 2383–2392, 2015.

[19] U. Kister, P. Reipschläger, F. Matulic, and R. Dachselt. Bodylenses - embodied magic lenses and personal territories for wall displays. In *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces*, page 117–126, 2015.

[20] S. Suzuki and K. Abe. Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing*, 30(1):32–46, 1985.

[21] K. Davila, F. Xu, S. Setlur, and V. Govindaraju. FCN-lecturenet: Extractive summarization of whiteboard and chalkboard lecture videos. *IEEE Access*, 9:104469–104484, 2021.