Construction of a Switching Support System for Live Broadcast of Oral Presentation

SHIORI ENDO^{1,a)} YOSHINARI TAKEGAWA¹ AYAKA FUNAKI¹ KOHEI MATSUMURA² KEIJI HIRATA¹
TAKEO IGARASHI³

Received: June 16, 2020, Accepted: December 1, 2020

Abstract: In this paper, we propose a switching support system for live broadcast of oral presentation. Analysis of the switching of professional switchers revealed that, in 57 oral presentations, professional switchers carry out switching based on events such as the presenter moving onto a new slide or pointing to the screen. In this paper, with the goal of even switching beginners being able to do switching intuitively, we proposed an event-based switching interface reflecting special characteristics of switching. We implemented a prototype of the proposed system and used it to conduct an evaluation experiment. The results of the experiment suggested that the proposed method can be operated as easily as the previous method and that the proposed method conforms to professional switching more than the previous method.

Keywords: Multi camera, GUI, Rule

1. Introduction

In recent years, the live transmission of research presentations via Ustream, Youtube etc. has become very popular. Since 2012, IPSJ (Information Processing Society of Japan) has been transmitting live broadcasts of speeches from the conferences it organizes, and other societies have similarly been transmitting live broadcasts. Live broadcast of conference speeches has the advantage that the results of research can be transmitted to not just the participants in the conference venue but also people not present. This allows the presenters to inform a larger number of people about their research. These live broadcasts are usually conducted in a multi-camera environment, in order to convey appropriate information to viewers. However, switching between multiple camera images (hereafter referred to as switching) appropriately is difficult.

Conferences are typically run by researchers, who may well have experience of acting as chairman, who have a great deal of presentation experience and who are familiar with the progression of presentations and research content. In live broadcasts at conferences, it is normally a researcher belonging to the conference who is responsible for switching, rather than a professional i.e. someone who has received specialized camera training or has filming work experience. Hereafter, we refer to a person responsible for switching as a switcher. The researchers are beginners regarding switching, not having received any specialized training.

Switchers need to select camera images in such a way that research content can be conveyed to viewers without inducing boredom, stress, or discomfort. On occasion, switchers also need to

respond flexibly to situations such as trouble occurring onstage or a cease in transmission due to content that infringes copyright.

We compared the average switch count in switching by professional and amateur switchers. The term ‘professional’ here refers to a person who has experience in switching as a job involving shooting and relaying. An ‘amateur’ is a person who has not received specialized education regarding switching, such as a steering committee member who belongs to an academic society. In conference oral presentations (IPSJ-ONE, which is a conference featuring oral presentations by well-known Japanese researchers related to IT), the average switch count in switching by the three professionals was 2.9 switches per minute. In contrast, the average count of the three amateur switchers was 0.1 switches per minute, giving us the analysis result that amateur switchers’ switch count is low at an academic symposium (IPSJ Interaction in 2016, 2017, and 2018).

As a result of interviewing 3 amateur switchers from an academic meeting (SIGMUS), we obtained comments such as ‘Switching is an operation that you have to stay focused on. Because it’s a lot of work for one person, we divide the work between several people and do it alternately’ and ‘It’s difficult to convey your know-how to students in a limited time, so the work can’t be left to student part-time workers and we’re chronically short of helpers’, showing that there is a problem relating to the difficulty of switching operation.

Accordingly, as a method to solve this problem, our research proposes a system to support switching by beginners.

In this research, to acquire switching know-how from professional switchers, we analyzed the switching professionals performed at IPSJ-ONE. From the results it was clear that the professionals performed switching based on events that occur during the presentations. Therefore, we designed and implemented an

¹ Future University Hakodate, Hakodate, Hokkaido 041–8655, Japan

² Ritsumeikan University, Kusatsu, Shiga 525–8577, Japan

³ The University of Tokyo, Bunkyo, Tokyo 113–0033, Japan

^{a)} g2119007@fun.ac.jp

event-based switching interface that reflects the special characteristics of switching by professional switchers. We conducted an evaluation experiment targeted at switching beginners, the results of which suggested that the proposed method can be operated as easily as the previous method and that the proposed method conforms to professional switching more than the previous method.

It is considered that, by using the proposed system, even switching beginners can deal with switching procedures, enabling the burden (live video broadcast) of management staff to be dispersed.

2. Related Work

There are several works on automatic camera switching for multi-view video.

Wirecast^{*1} and vMix^{*2} are very popular software video switcher commercially available. These have a simple feature for automatic switching. The feature allows users to create a playlist and automatically rotates multiple cameras according to the playlist. Open Broadcaster Software (OBS)^{*3} is also a prevalent opensource software for live streaming. Thanks to contributors to the opensource project, OBS has a plug-in feature. Plug-ins include expanding OBS functionality such as automatic scene switching for a live game streaming that automatically detects an active window on a computer and switches source for broadcasting to the active window. Although these features are functional for limited use, they cannot recognize scenes nor switch to an appropriate source.

Wang et al. proposed an automatic camera switching system for soccer. They calculated scores for each camera based on the presence of a ball or player [14], [15], trajectory of viewing target [17], [18] and interest of viewer groups [16]. Tang and Boring developed a video highlighting system [11]. They used timeline stream of sports fans on social media (e.g., Twitter) to semantically annotate live broadcast sports games. Leake et al. proposed a system for editing video of dialogue-driven scenes [6]. The system automatically selects the most appropriate clip from multiple cameras based on a user-specified set of film-editing idioms. The work of Sugimoto et al. [20] generates realistic lecture videos containing a lot of information necessary for students in real time, by selecting images according to switching rules generated in advance. Onishi et al. [7] proposed a system for producing video images for distance learning based on distributed information. Above systems mainly rest on automatic activity recognition and understanding on video scenes and it is still a challenging task. So, we would rather involve human at the first step, then use data obtained from the first step to design an automatic video switching.

There are also a number of researches on video editing utilizing human resources including crowdsourcing. Bernstein et al. proposed a framework to be able to realize on-demand and real-time crowdsourcing [1]. As an example of their framework, they demonstrated a crowd-powered camera where crowd workers create the best single moment from video footage. Truong

et al. present an interactive video editing tool with voiceover function [12]. The tool takes an audio recording of the narration, automatically transcribes it and then aligns a transcript of the narration with video footages. Using the above information the tool enables authors to quickly seek story events in the narration with video footage. VidCrit is one of interesting attempt to review a video in the process of editing [10]. VidCrit captures reviewers' spoken comments and mouse interactions as well as physical actions of the reviewers and provides feedback interfaces of transcribed/segmented reviews. Craggs et al. presented a query-sensitive animated preview generated from crowdsourced semantic tagging [3] for online videos. DeCamp and Roy proposed a system for annotating the location and identity of people and objects in a multi-camera video [4]. This enables human-machine collaborative systems that achieve a greater accuracy than machine-only approaches at less cost than human. Most of the above researches used human-resources for better performances on complicated tasks. However, even for a human, it is not easy to do tasks processes in real-time. Tasks that require real-time processing such as multi-view video switching need to be processed by many humans, with computing power, or broken down into smaller tasks. Our approach is providing the event-based switching interface to break down a complicated switching task into a small task.

Our interface also works as a real-time annotation tool for multi-camera presentation videos. Annotation tools is a still an attracting theme of video researches. Weher and Poon developed a pen-based video annotation tool that enables users to correlate their notes and keywords with a videotape during recording [19]. Also, annotation tools for a specific category video are proposed. For example, Utasi and Benedek proposed a tool for people detection [13], Cabral et al. [2] is for contemporary dance, Miller et al. is for a moving target [8] and how-to videos [9]. To our knowledge, there are not many annotation tools for multi-camera video, especially for research presentations. We attempt to clarify the characteristics of research presentation recorded with multiple cameras. This contributes to developing the tools not only for video footage of research presentations but also for real-time, multi-viewed video.

3. Switching Analysis

This section consists of two subsections. To design a switching support system for non-professionals, we need to clarify the burdens that non-professional (i.e., amateur) switchers feel. We investigate this through the interview with the amateur switchers, described in the first Section 3.1. In the second Section 3.2, we also investigate how professional switchers perform their switching by analyzing professional switching work at a series of academic conferences.

3.1 Amateur Switcher

In the oral presentation section of an academic meeting (SIGMUS), amateur switchers perform switching and the resulting video is transmitted to YouTubeLive. Switching and relay work is carried out by one person. The relay video of oral presentations transmitted at SIGMUS is a typical example of a

^{*1} <https://www.telestream.net/wirecast/>

^{*2} <https://www.vmix.com>

^{*3} <https://obsproject.com>

live broadcast at an oral presentation, which is the assumed scenario in this research. We analyzed the transmitted video from SIGMUS and conducted in-depth interviews with the amateur switchers. These interviews were conducted to clarify what kind of burden amateurs usually feel and what they find difficult, and to think about specific support methods. In interviews with amateur switchers, they expressed the opinion that they have no knowledge of live broadcasting and switching, and even if they try to switch in a way that the viewer wants to see, in the end they basically do not switch because they do not know what viewers need and at what time they need it. It was also found that it is difficult to teach a person without knowledge how to use a device for live broadcasting in a short time, because a certain amount of knowledge is required to operate it. Therefore, we thought that it would be necessary to provide support that makes it easy to learn how to use the system and allows switching without having to think about which camera image to project. The details of the interview are described in Section 3.1.1.

We analyzed the number of switches, as well as the corresponding trends, of the 47 pieces of oral presentation videos on which switching had been performed at SIGMUS. The average number of switches in the amateur switching was 0.1 switches/minute. Switching was extremely mechanical, consisting only of selecting the image of the presenter immediately after the start of a presentation, selecting the image of slides during the presentation, stopping transmission of NG material, and switching to the image of the questioner during the Q&A session.

We conducted in-depth interviews with the three amateur switchers responsible for image switching at SIGMUS. The details of the in-depth interview are given below.

Interview time

The interview time was approximately one hour per interviewee.

Number of interviewers

Interviews were conducted by two of the authors of this paper.

Interview record

Interviews were recorded with a video camera. The consent of the interviewees to have their interviews recorded was obtained in advance.

Consideration of questions

Three of the authors considered the question categories before the interviews were conducted. Since the purpose of the interview is to clarify the burden of the amateur switcher, we categorized the three view-points of technical, mental, and physical difficulties as the following questions.

- What led you to be placed in charge of live broadcast work?
- Is live broadcast work a big responsibility?
- What do you take particular care over in your live broadcast work?

These categories were considered based on observation of the interviewees when they performed switching. The interview was semi-structured, that is, we initially started the interview with the above three questions, then proceeded to inquire into the subjects' inner thoughts, such as the reasons and contexts for their answers.

3.1.1 Interview Results

The utterances of the interviewers and interviewees were tran-

scribed and analyzed using subject analysis. Transcription and analysis were carried out by three of the authors of this paper.

Work responsibility in a state of having no specialist knowledge

The following comments reveal that the interviewee felt uncertainty due to lack of specialist knowledge of live broadcast and switching.

"Roles are assigned during the conference's management committee meeting. I was asked to be in charge of live broadcast, but I don't have any specialist knowledge."

"I just take care to ensure that there is no deterioration in basic video quality, for example by checking there is no distortion of the audio, or that the video doesn't cut out part way through."

"I want apply switching in a way that will make viewers want to watch, but, because I don't know what image viewers require and when, I basically try not to switch."

Size of work responsibility

From the above comments, it became clear that relay work imposes a heavy burden.

"It is a big responsibility for one person, as the work continues from morning to evening, so we try to alternate as much as possible."

"It is one of the biggest burdens among the work of the management committee. Originally, we wanted to hire an outsider, but due to budgetary problems we are unable to do so."

"Because conference management involves many jobs besides live broadcasting, the current three people assigned to relay is the limit, although we would like to be assigned more people."

"Setting up the equipment is complicated, because the environment differs with each venue."

"The members in charge of relay go to the venue and set everything up. As it is necessary to do this either the day before or on the morning of the conference, there are times when we only set up the equipment then go home without participating in the conference."

"I can't relax during the presentations. If my concentration wanes, I miss my timing to switch."

"Participating in all five of the conferences held each year is a big burden."

Difficulty of dividing up the work burden

From the comment below, the difficulty of dividing the work burden is evident.

"Reception work can be carried out as a part-time job by students, but operating live broadcast equipment requires certain knowledge. Teaching this in a short time to students with absolutely no knowledge is difficult."

It is clear that, in switching by amateur switchers, while there is the desire to transmit content that is valid for the viewers, there is also a state of negativity, whereby fear of making a mistake results in dull, uniform switching.

3.2 Professional Switcher

We analyzed the switching performed by professional switchers in 3 years' worth (the past 3 years from 2015 to 2017) of oral presentation video from IPSJ-ONE, a conference organized by IPSJ. Each presentation was approximately 5 minutes long

Fig. 1 Camera view.

and there were 19 presenters in 2015, 18 in 2016 and 20 in 2017, making a total of 57. Professional switchers have over 10 years of practical experience performing switching in TV or internet live broadcasting. To the best of our knowledge, there was no given manual or methodology relating to switching; the professional switchers conveyed their switching know-how to their underclassmen (apprentices) verbally. This has been confirmed by a professional switcher who actually carried out switching at IPSJ-ONE. Thus, to investigate switching support methods and functions necessary for switching support, we analyze switching by professional switchers.

Depending on the switcher, there are differences regarding individuality, functions that are used, and functions that are not used. This data is just one example of professional switching, not a representative example of all professional switching. In this paper, we discuss the data extracted from the common features of the three professional switchers, which has a certain validity.

The coding [5], categorization, and analysis described hereafter were conducted by three of the authors while watching archived oral presentation footage from IPSJ-ONE. The participants viewed all the videos and implemented open coding. Next, the members confirmed with each other that the labels and contents matched, then implemented focused coding to scrutinize the label names. In addition, we interviewed a professional switcher responsible for oral presentation switching at IPSJ-ONE and had him confirm the analysis results. The facts demonstrating that the analysis results and interview results were in agreement are presented below.

3.2.1 Camera View

As shown in **Fig. 1**, regarding the image ultimately seen by viewers (hereafter referred to as final image) we observed four types of camera view: ‘presenter’, ‘slide’, ‘entire stage’, and ‘picture-in-picture (a screen composed of slide (main screen) and presenter (sub-screen)).’ Furthermore, the position of the horizontally oriented sub screen in picture-in-picture is decided in accordance with the relationship between the positions of the slide and lectern.

3.2.2 Switching Frequency

We investigated the frequency of switching by professional switchers in each year.

Average value and standard deviation of switching count

The average switching count of the 57 presentations made from 2015 to 2017 was 14.5 switches per presentation (SD 4.5). The

Fig. 2 Average camera view continuity time per scene.

Fig. 3 Number of appearances of each camera view per presentation.

average switching count per presentation was approximately the same in each year, with little variation.

Appearance frequency and continuity time of each camera view

The average value and standard deviation of continuity time are shown in **Fig. 2**. Continuity time refers to the time from one camera view being selected as the final image until it switches to a different camera view final image. For all camera views there are almost no cases of switching in one or two seconds and mostly switching is not performed until five seconds after the previous switch. This is because performing switches in too short a time gives viewers the impression that a switching error has been made, which leads to a decrease in the quality of video.

The average of the frequency of appearance of each camera view is expressed in **Fig. 3**. The frequency of appearance of ‘Picture in picture’ is low compared to the other three. However, looking at the continuity times shown in **Fig. 2**, it can be seen that this camera view is not totally unused.

From **Fig. 2**, it is evident that ‘slide’ was being selected for a longer time than any other camera view. As oral presentations have the particular characteristic of using slides, ‘slide’ being selected for a long time can be considered valid.

However, regarding continuity time and frequency, views of the presenter alone, the entire presentation (presenter and slides) and picture-in-picture were being selected the same amount as slides. From this we can expect the effect that, by showing the entire venue, the atmosphere in the venue can be conveyed.

Fig. 4 Switching transition pattern rate per presentation.

Fig. 5 Conditions for switching during presentation.

Furthermore, another reason for selecting views including the presenter is that there are presenters who prefer the presentation style of explaining verbally rather than putting a large amount of text on their slides. Also, there are presenters who give supplementary explanation of the content of slides by skillfully using gestures, hand motions and facial expressions.

3.2.3 Switching Conditions

The average of the switching transition pattern rate for the three years are shown in Fig. 4. Switching transition pattern refers to the bigram of camera views arranged in chronological order. Figure 5 presents the results of surmising and organizing the bases for switching, such as slide content and presenter's speech or gestures, for all the switching that occurred during oral presentations. We hereafter explain each condition.

As our research is targeted at beginners and we want to make switching simple, in the switching transition pattern rate of each year we excluded patterns with a rate of below 5% from the transition candidates. That this condition is valid was checked with one of the professional switchers responsible for switching at IPSJ-ONE.

Conditions for Selecting Image of Presenter

The cases of an image of the presenter being selected are primarily the times when the presenter is explaining about something other than the information displayed on the slide. As a typical example, when only a concept or keyword is displayed on the slide and the presenter is verbally giving additional explanation of the intention or content, after an image of the slide has been selected it is switched to an image of the presenter.

Furthermore, in the case of the presenter using a lot of gestures or explaining using the actual machine they have developed, image of the presenter is selected.

Conditions for Selecting Entire Presentation Image

The cases of entire presentation image being selected are those in which the presenter is explaining while pointing to or looking towards the screen.

Showing both the presenter and screen to the audience makes it easy to understand what part the presenter is trying to explain.

Also, entire presentation image was often selected when a demo video had just finished. This is because when the video finishes the presenter usually verbally gives thoughts or elucidations regarding the video, while using the screen to explain. Accordingly, the entire presentation image, showing both the screen and presenter, is appropriate.

Conditions for Selecting Slide Image

The cases of slide image being selected were the times when the presenter switched slides or began to play a demo video. This is because, in these instances, what, for the viewer, is new information is displayed on the screen.

Additionally, when camera image of the presenter has been selected but the presenter starts to re-explain content from the slides, or when camera image of the entire presentation has been selected but the slide content is difficult to see, camera image of slides is selected.

Conditions for Selecting Picture-in-picture Image

According to the results of analysis of switching transition pattern, the image transitions from slide image to picture-in-picture image, then back to slide image.

In slides displayed on the main screen, when the presenter is explaining a part not contained within the picture-in-picture superimposition area shown in Fig. 5, the image transitions to picture-in-picture image, then when the presenter is explaining a part within the superimposition area the image returns to slide.

In this paper, when the part the presenter is explaining is within the picture-in-picture superimposition area and the slide on the main screen is hidden by the sub-screen, we describe this as having picture-in-picture overlap.

4. Switching Support System Design

Our system is supposed to be used in live broadcast of oral presentations featuring slides. In switching of oral presentations containing slides, it is necessary to consider slide content as well as the presenter's speech, gestures, gaze and the direction he or she is facing. Nevertheless, as mentioned at the beginning of this paper, the actual presentation location is a place where complex factors are intertwined. For this reason, switching beginners do not know what camera image they should select in which situation. From the results of analyzing switching by professional switchers, we understood that professional switchers perform switching based on events that occur during oral presentations. Therefore, in this research we make events into objects, in the form of buttons, and propose a GUI featuring a button corresponding to each event. When a button is pressed the system recognizes that an event has occurred and then, from a combination of this event and the current camera status, the system automatically selects

Fig. 6 Screen snapshot of proposed user interface.

the next camera image to transition to. With this system, beginner switchers only have to observe events that occur during a presentation and do not have to think about what camera image they should select.

All the functions of the proposed system utilize the knowledge related to camera view and frequency mentioned in Section 3.

4.1 System Structure

In the case in which three cameras are set up in the presentation venue, one camera records the presenter, one records the screen and one records the entire presentation. The system obtains the image from each camera and switchers use the GUI provided by the system to select camera images. Then, using a transmission device, the final image is made available on a live broadcast website, such as NicoNico Live Broadcast.

4.2 User Interface

We explain the functions of the system that uses the user interface, with reference to Fig. 6. Incidentally, the roman numerals in Fig. 6 correspond to the roman numerals used in this section.

(i) Final Image

This function presents the final image that will be seen by viewers. From this, switchers can confirm with what kind of image viewers will ultimately be provided.

(ii) Entire Stage Image

In contrast, displaying a large image of the entire presentation enables switchers intuitively to notice events that occur during verbal presentations, such as the conditions around the stage and the presenter's gestures.

Also, it is essential to confirm whether a picture-in-picture overlap is occurring when slide image is selected. In this case, on the area of the final image a dotted line rectangle is displayed to represent the sub-screen area. Using this, switchers can confirm whether or not there is a picture-in-picture overlap.

(iii) Switching Continuity Time

This function presents switching continuity time. When the image switches to a different camera view, continuity time is reset. As was mentioned in Section 3.2.2, repeatedly performing

switching in a short time gives viewers the impression that there has been a switching operation error, which lowers the quality of the video. For this reason, we provide this function so that switchers will be aware of continuity time and check it.

(iv) Average Switching Count

We provide a function for confirming the average switching count per minute. With this function switchers can check that switching is not becoming fixed.

(v) Event Buttons

This is a function by which the optimum camera image is automatically switched to when the switcher selects the event button corresponding to the presenter's event. We made the switching conditions explained in Section 3.2.3 into objects, in the form of event buttons, and configured them. Buttons corresponding to the conditions in Fig. 5, which are not dependent on pre-transition camera view (except video playback), have a high usage frequency and are thus larger than the buttons corresponding to the conditions in Fig. 5, which are dependent on pre-transition camera view. In addition, event buttons that can only be used alternately, such as 'film playback start–film playback end', 'with picture-in-picture overlap–without picture-in-picture overlap', are toggle buttons, meaning that the title written on the button changes each time it is pressed.

Furthermore, besides the conditions explained in Section 3.2.3, there is also an 'NG' event button. In live broadcast of academic conferences, there are cases in which parts of presentations cannot be broadcast, due to circumstances relating to copyright or special permission. This kind of non-broadcastable (no good: NG) content can be generally classified into three types: NG image, NG audio and all NG (image + audio). Pressing the NG event button causes the NG related event buttons (NG image, NG audio, all NG) to appear, and, by selecting the appropriate one of these buttons, transmission of part, or all, of the broadcast can be stopped.

(vi) Presentation Progression Control Button

Due to limitations of space we cannot provide detailed explanation in this paper, yet it should be mentioned that the proposed system considers not only switching during presentations, but

also in situations where an emcee or questioners appear, such as during presentation preparation or question and answer sessions.

Pressing the presentation progression control button reloads the camera image and event buttons necessary at the relevant stage of the presentation. Switching can be performed by the same method as during a presentation, by pressing event buttons.

5. User Study

We conducted an evaluative experiment to investigate the usefulness of the proposed system.

Procedure

We compared our proposed method and a previous method. As shown in **Fig. 7**, in the previous method event buttons are made unavailable and switching is performed by clicking on thumbnail images. In both methods, we did not display auxiliary information such as current time, elapsed switching time, average switch count, presentation progression management button and NG button. This is because we wanted to investigate the effect of event buttons in this experiment.

An experimental task was carried out in which subjects performed switching on one ten-minute presentation from a domestic workshop (WISS 2017). In the experiment we had subjects view the presentation using a 13.1 display and earphones, and use a mouse to select buttons or thumbnail images.

The subjects' selected buttons or thumbnail images and data was obtained.

We created the correct answer data in advance, by having the professional switcher mentioned in Section 3.2 use the previous method to perform switching on the same ten-minute presentation, which produced a 'correct answer video' of the presentation. Furthermore, in order to evaluate the correct answer video we also created a randomly switched video and a video without switching, and carried out an evaluation of the three types of video. It was confirmed that the video of the correct answer data had the highest evaluation regarding video quality. It is difficult to define uniquely what is a good video for all viewers, because viewers are diverse. Nevertheless, professional videos are relatively popular with various people. Therefore, in this project, the goal is to be able to switch in a similar way to a professional, by using the system.

In this experimental task, the consistency rate of the subjects' data with the correct answer data was calculated. The correct answer data is in chronological order as with the subjects' data. The method of calculating the consistency rate is explained below. At first, when comparing the two sets of data, the value is set to 1 if the camera image selected at a certain time in the subject data is the same as the camera image selected at that time in the correct answer data. If the images are not the same, the value is 0. After setting 1 or 0 for every frame of the video, the total value is calculated. The consistency rate is the result of dividing the total value by the total number of seconds of the video. The higher the consistency rate is, the closer switching is to the correct answer, in other words the more it can be said to conform to the switching of professionals.

As subjects we took 24 university students with no specialist knowledge relating to film or filming. We assigned 12 subjects

Fig. 7 Screen snapshot of previous user interface.

Table 1 Results.

Metric	Previous method		Proposed method	
	Ave.	SD	Ave.	SD
Consistency rate with correct answer data	49.5	12.2	71.0	8.55
Switching frequency (switches per minute on average)	1.41	8.59	3.70	6.27
How easy was the system to operate?	4.17	0.37	4.25	0.43

to each method, making the study a between-subject study. Also, the subjects had no prior knowledge of the oral presentation on which they were asked to perform switching, and they saw the presentation footage for the first time in the experiment. The experimenters explained to the subjects, in advance, how to operate in each method, and the subjects practiced until they understood or for up to ten minutes. After the experiment, we had subjects answer a written questionnaire, shown in **Table 1**, based on the five level Likert Scale (1: difficult–5: easy) and explain freely the reasons for their answers.

Results

Table 1 presents the average of the consistency rate, the average of switching frequency and the average of ease of use of the system of each method. In addition, the standard deviation of each method is also presented in the same table. When the Welch's t-test was applied at the significance level of 1%, a significant difference was observed between previous method and proposed method ($t(19) = 4.78, p < .01$). Regarding the switching frequency, the correct answer data was 3.6 switches per minute on average, while the previous method was 1.4 switches per minute on average and the proposed method 3.7 switches per minute on average. As a result, the proposed method is closer to a professional in terms of switching frequency. When Welch's t-test was applied at the significance level of 1%, a significant difference was observed between the previous method and proposed method ($t(19) = 4.56, p < .01$). Focusing on the questionnaire results, it can be seen that, regarding the simplicity of the operation, both the previous method and proposed method received similar ratings. As a result of applying the Mann-Whitney U test, a significant difference was observed ($U = 3.00, p < .01$) between the previous method ($M = 4.17, MD = 4.00, SD = 0.37$) and proposed method ($M = 4.25, MD = 4.00, SD = 0.43$).

Consideration

The subjects who used the previous method performed switching focusing on the image they wanted to see. As a result, the consistency rate was low because this was different for each subject.

There was the comment from among the six subjects who used the previous method: 'I selected the image I wanted to see, but I

didn't know if that was the image other people would want to see. So I wasn't sure if I had chosen the correct image.' This shows that their lack of switching knowledge and experience meant the subjects were not able to make decisions with confidence. There was no such comment from the subjects who used the proposed method.

We created a professional switching rule in advance and designed it so that a suitable camera can be selected by inputting the status of the current presentation. Accordingly, we think that the subject can operate the system like a professional. To switch the switching source (camera image), it is necessary to judge the situation and select an appropriate source. In addition to judging by looking at the status of the presentation, it is also necessary to understand the interface used for switching, such as the correspondence between buttons and camera images. Even when determining an event, experience and knowledge are important in determining which event is key. In the proposed method, information is given that limits which event to see from among many events. This system allows switching by performing simple tasks that anyone can do.

On the other hand, a subject who used the proposed method commented: 'it was difficult to judge whether a certain action qualified as a gesture or pointing.' Evidently, as there are event buttons that require the switcher to interpret their meanings for him or herself, there are situations in which a clear judgement cannot be made. Although there was uncertainty and hesitancy regarding discrimination of events, the problem of the proposed method is more substantiated than in the previous method. We consider that the proposed method contributes to the simplification of switching operation.

The target value of the consistency rate with the correct answer data has not been clearly set at this time. Calculation of the required consistency rate will be an issue for the future. As for how the target value will be calculated, we will calculate the consistency rate of switching data between multiple professional switchers, or have them watch multiple videos with different consistency rates from the correct answer data and give a score to each video, and so on. By such an experiment, the target value of the consistency rate can be calculated.

Regarding the switching frequency, the proposed method was closer to the switching frequency of the correct answer data than the previous method, but both methods had a high switching frequency overall. The cause of this is considered to be that the video of the presentation used for the user study was a video in which switching is likely to occur frequently in the first place.

6. Conclusion

In this research we constructed a switching support system for live broadcast of oral presentations. We analyzed the trends and conditions of switching by professional switchers and constructed an event-driven interface based on the analysis results. The evaluative experiment suggested that the proposed method can be operated as easily as the previous method and that the proposed method conforms more closely to switching by professional switchers than the previous method.

As a future work, we plan to evaluate oral presentations with

varying fields of technology and times, and presentations including NG broadcast. We also plan to evaluate the quality of video created by beginner switchers, as well as other functions, such as switching continuity time, and carry out practical application of our system. In the future, we plan to aim for an automatic switching system, however, to achieve this we must first overcome the problem of interpreting the meaning of each kind of event.

References

- [1] Bernstein, M.S., Brandt, J., Miller, R.C. and Karger, D.R.: Crowds in Two Seconds: Enabling Realtime Crowd-Powered Interfaces, *Proc. 24th Annual ACM Symposium on User Interface Software and Technology, UIST '11*, pp.33–42, Association for Computing Machinery (online), DOI: 10.1145/2047196.2047201 (2011).
- [2] Cabral, D., Carvalho, U., Silva, J.A., Valente, J.A., Fernandes, C. and Correia, N.: Multimodal Video Annotation for Contemporary Dance Creation, *CHI '11 Extended Abstracts on Human Factors in Computing Systems, CHI EA '11*, pp.2293–2298, Association for Computing Machinery (online), DOI: 10.1145/1979742.1979930 (2011).
- [3] Craggs, B., Kilgallon Scott, M. and Alexander, J.: ThumbReels: Query-sensitive web video previews based on temporal, crowd-sourced, semantic tagging (online), DOI: 10.1145/2556288.2557249 (2014).
- [4] DeCamp, P. and Roy, D.: A Human-Machine Collaborative Approach to Tracking Human Movement in Multi-Camera Video, *Proc. ACM International Conference on Image and Video Retrieval, CIVR '09*, Association for Computing Machinery (online), DOI: 10.1145/1646396.1646436 (2009).
- [5] Ezzy, D. and Ezzy, T.: *Qualitative Analysis: Practice and Innovation*, Social research today, Routledge (2002).
- [6] Leake, M., Davis, A., Truong, A. and Agrawala, M.: Computational Video Editing for Dialogue-Driven Scenes, *ACM Trans. Graph.*, Vol.36, No.4 (online), DOI: 10.1145/3072959.3073653 (2017).
- [7] Onishi, M., Izumi, M. and Fukunaga, K.: Automatic Production of Video Images for Distance Learning System Based on Distributed Information, *IEICE Trans. Information and Systems*, Vol.82, No.10, pp.1590–1597 (1999) (online), available from (<https://ci.nii.ac.jp/naid/110003183529/>).
- [8] Miller, G., Fels, S., Al Hajri, A., Ilich, M., Foley-Fisher, Z., Fernandez, M. and Jang, D.: MediaDiver: Viewing and annotating multi-view video, *CHI EA 2011 - 29th Annual CHI Conference on Human Factors in Computing Systems, Conference Proc. and Extended Abstracts*, pp.1141–1146 (online), DOI: 10.1145/1979742.1979711 (2011).
- [9] Nguyen, P., Kim, J. and Miller, R.C.: Generating Annotations for How-to Videos Using Crowdsourcing, *CHI '13 Extended Abstracts on Human Factors in Computing Systems, CHI EA '13*, pp.835–840, Association for Computing Machinery (online), DOI: 10.1145/2468356.2468506 (2013).
- [10] Pavel, A., Goldman, D.B., Hartmann, B. and Agrawala, M.: VidCrit: Video-Based Asynchronous Video Review, *Proc. 29th Annual Symposium on User Interface Software and Technology, UIST '16*, Association for Computing Machinery, pp.517–528 (online), DOI: 10.1145/2984511.2984552 (2016).
- [11] Tang, A. and Boring, S.: EpicPlay: Crowd-Sourcing Sports Video Highlights, *Proc. SIGCHI Conference on Human Factors in Computing Systems, CHI '12*, pp.1569–1572, Association for Computing Machinery (online), DOI: 10.1145/2207676.2208622 (2012).
- [12] Truong, A., Berthouzoz, F., Li, W. and Agrawala, M.: QuickCut: An Interactive Tool for Editing Narrated Video, *Proc. 29th Annual Symposium on User Interface Software and Technology, UIST '16*, pp.497–507, Association for Computing Machinery (online), DOI: 10.1145/2984511.2984569 (2016).
- [13] Utasi, A. and Benedek, C.: A Multi-View Annotation Tool for People Detection Evaluation, *Proc. 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications, VIGTA '12*, Association for Computing Machinery (online), DOI: 10.1145/2304496.2304499 (2012).
- [14] Wang, X., Hirayama, T. and Mase, K.: Viewpoint Sequence Recommendation Based on Contextual Information for Multiview Video, *IEEE MultiMedia*, Vol.22, No.4, pp.40–50 (online), DOI: 10.1109/MMUL.2015.75 (2015).
- [15] Wang, X., Muramatsu, Y., Hirayama, T. and Mase, K.: Context-Dependent Viewpoint Sequence Recommendation System for Multiview Video, *2014 IEEE International Symposium on Multimedia*, pp.195–202 (online), DOI: 10.1109/ISM.2014.44 (2014).
- [16] Wang, X., Enokibori, Y., Hirayama, T., Hara, K. and Mase, K.:

User Group Based Viewpoint Recommendation Using User Attributes for Multiview Videos, *Proc. Workshop on Multimodal Understanding of Social, Affective and Subjective Attributes, MUSA2 '17*, pp.3–9, Association for Computing Machinery (online), DOI: 10.1145/3132515.3132523 (2017).

- [17] Wang, X., Hara, K., Enokibori, Y., Hirayama, T. and Mase, K.: Personal Multi-View Viewpoint Recommendation Based on Trajectory Distribution of the Viewing Target, *Proc. 24th ACM International Conference on Multimedia, MM '16*, Association for Computing Machinery, pp.471–475 (online), DOI: 10.1145/2964284.2967265 (2016).
- [18] Wang, X., Hara, K., Enokibori, Y., Hirayama, T. and Mase, K.: Personal Viewpoint Navigation Based on Object Trajectory Distribution for Multi-View Videos, *IEICE Trans. Information and Systems*, Vol.E101.D, No.1, pp.193–204 (online), DOI: 10.1587/transinf.2017EDP7122 (2018).
- [19] Weher, K. and Poon, A.: Marquee: A Tool for Real-Time Video Logging, *Proc. SIGCHI Conference on Human Factors in Computing Systems, CHI '94*, pp.58–64, Association for Computing Machinery (online), DOI: 10.1145/191666.191697 (1994).
- [20] Sugimoto, Y., Marutani, T., Kakusho, K. and Minoh, M.: Lecture Context Recognition based on Statistical Feature of Lecturer Action for Automatic Video Recording, Vol.153, pp.179–186 (2006) (online), available from (<https://ci.nii.ac.jp/naid/110004788342/>).

Shiori Endo graduated from the Department of Information Architecture, Faculty of Systems and Information Sciences, Future University Hakodate in 2019. She is currently enrolled in the Graduate School of Systems and Information Sciences at the same university. Her research interests

include switching between multiple camera images.

Yoshinari Takegawa is an Associate Professor at Faculty of Systems Information Science, Future University Hakodate. He received his B.Eng. from Mie University in 2003, and received his MEng and Ph.D. from Osaka University in 2005 and 2007, respectively. His research interests

include human-computer interaction, educational technology, and entertainment computing.

Ayaka Funaki graduated from the Department of Information Architecture, Faculty of Systems and Information Sciences, Future University Hakodate in 2018. Her research interests include switching between multiple camera images.

Kohei Matsumura is an Associate Professor at Ritsumeikan University. He received his Ph.D. in Knowledge Science from Japan Advanced Institute of Science and Technology in 2010. His broad research interests include human-computer interaction, interaction design, urban informatics, and embodied cognitive science.

He is a member of the IPSJ and ACM.

Keiji Hirata Doctor of Engineering from University of Tokyo in 1987. Joined NTT Basic Research Laboratories in 1987 (later changed to NTT Communication Science Laboratories) and Future University Hakodate as professor in 2011 (–present). His research interest includes computational music theory, algebraic framework for music creation, mobility as a service (MaaS), and multimodal learning support system.

Takeo Igarashi Professor, Department of Computer Science, The University of Tokyo. Dr. Igarashi received a Ph.D. from the Department of Information Engineering at The University of Tokyo in 2000. He joined the Department of Computer Science as an assistant professor in 2002 and was promoted to a professor in 2011.

His research interest is in user interfaces and interactive computer graphics in general. He is known for the development of a sketch-based modeling system (Teddy) and a multi-touch shape deformation method. He led a JST ERATO project and a JST CREST project as a director. He received the SIGGRAPH Significant New Researcher Award, JSPS Award, and CHI Academy Award.