Improvement of Presence in Live Music Videos and Alleviation of Discomfort of Viewers by Zooming Operation

Ai Oishi *, Eiji Kamioka, Phan Xuan Tan, and Manami Kanamaru

Graduate School of Engineering and Science, Shibaura Institute of Technology, Tokyo, Japan;
Email: {kamioka, tanpx, kanamaru.manami.a4}@shibaura-it.ac.jp
*Correspondence: ma21024@shibaura-it.ac.jp (A.O.)

***Abstract*—People can enjoy watching live performances without visiting live venues thanks to the development of live music streaming services on the Internet. However, such live music videos, especially those recorded by amateur band members, lack a sense of presence. Therefore, in the previous study, the authors proposed a method to improve the sense of presence in live music videos by performing zooming on the video frames. It achieved enhancing the sense of presence. However, it also increased the discomfort of the viewers. This is because the zooming was performed not on a music performer but in the center of the screen, resulting in an unnatural experience for the viewer. Therefore, in this paper, a new zooming method, which effectively emphasizes the music performer with intense movement, is proposed, introducing the concept of the "Main Spot". The evaluation results through an experiment verified that the proposed method improved the sense of presence in live music videos and alleviated the discomfort of the viewers.**

***Keywords*—presence, main spot, Region of Interest (ROI), peripheral vision, live music video**

I. INTRODUCTION

The rapid development of online multimedia streaming services, e.g., YouTube, greatly allows the users to enjoy various recorded live music videos uploaded by not only professional music bands but also amateur ones. For such a service, it is necessary to enhance the "presence" [1, 2] in each video. Thereby, the fans will have the feeling of "being there" with their idols. Professional music bands can conduct the postproduction step pretty well with the aid of a professional camera crew and video editing team. However, with the limited supportive devices and human resources, the products of amateur bands usually lack a sense of presence. Here, the sense of presence felt by humans is defined as "feeling as if they were in the actual place". The presence is composed of multiple sensory components [3], which are the spatial, temporal, and physical components [4]. "Activity", one of the factors of the temporal component, is proved to fluence the sense of presence [5]. The "activity" refers to a sense of change or a sense of motion, which is perceived by the peripheral vision in human vision systems [6]. Some studies [7−9] achieved improving the sense of presence by stimulating peripheral vision with optical illusions [10, 11]. However, since the optical illusions are displayed by adding external effects to the original image on the screen, they are not versatile and the image size is limited.

In the previous study, zooming operations in scenes with the intense movement were proposed to avoid the use of external effects for activating peripheral vision to improve the sense of presence in live music videos [12]. Concretely, a zooming operation was gradually performed on continuous video frames when a large optical flow in a video scene was detected. The evaluation experimental results showed that the viewers felt an improvement in the sense of presence. However, they also felt discomfort in some situations. This is because the zooming operation was simply performed in the center of the video frame. In this case, if music performers are not in the center of the screen, the zoomed video frames are not natural, resulting in making the viewers feel discomfort. This leads to a hypothesis that if a music performer who is moving most intensely is detected and the zooming is performed around the music performer, the sense of the presence of the video can be improved while alleviating the discomfort of the viewer.

The purpose of this study is to alleviate the discomfort of the viewers as well as to improve the sense of presence while watching live music videos. To this end, in this paper, a new video zooming method is proposed, which firstly identifies the region of interest called "Main Spot" in the video frame, and then the average optical flow value in the detected Main Spot is calculated and compared with the pre-defined threshold. Note that in this paper, "average optical flow value" means the average value of the magnitude of the optical vectors detected in a given region. Once the average optical flow value of the Main Spot in a video frame exceeds the threshold, the zooming operation will be triggered to enlarge the Main Spot in the continuous video frames. To realize the proposed method, three things must be clarified as follows:

Manuscript received August 15, 2022; revised September 20, 2022; accepted October 19, 2022.

doi: 10.18178/joig.11.3.233-241

(1) How can music performers be detected?
(2) How can the Main Spot be identified among the detected music performers?
(3) How can the threshold of the average optical flow value be determined for the trigger of zooming for the Main Spot?

In this paper, the above requirements will be thoroughly discussed through some preliminary experiments. In addition, the results of the evaluation experiment will demonstrate that the sense of presence is improved and discomfort to the viewers is alleviated.

The rest of this paper is organized as follows: Related work is described in Section II. The proposed zooming operations are stated in Section III. Then, the evaluation and the discussion are given in Section IV and Section V, respectively. Finally, Section VI concludes this paper.

II. Related Work

A. Peripheral Vision

According to Miura [13], the human visual field range is defined as shown in Fig. 1. The area of the human visual field within which a person can clearly perceive objects is called "useful field of view", and the visual fields other than "central visual field" including the useful field of view are called "peripheral visual fields". The central visual field is sensitive to color and insensitive to movement, while the opposite is found in the peripheral visual field [14]. This indicates that generating stimuli such as optical illusions as optical flow in the peripheral vision evokes a sense of "I am moving", in other words, a sense of activity in humans [4]. On the other hand, the sense of activity strongly correlates with the sense of presence [5]. Thus, the sense of presence can be improved by stimulating peripheral vision and enhancing the sense of activity.

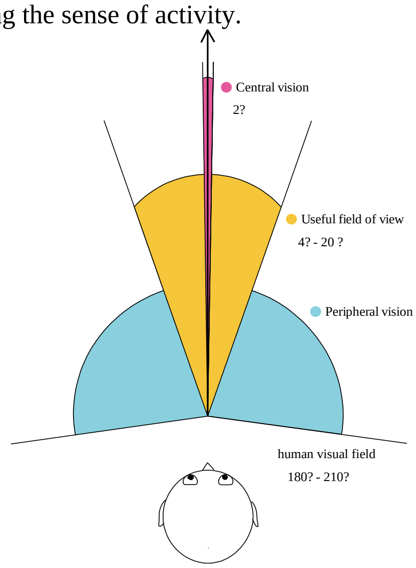


Figure 1. Range of human vison.

B. Optical Illusions

There exist a lot of studies on visual illusions to stimulate peripheral vision. Jones *et al.* [7, 8] developed "IllumiRoom" using visual illusions. They projected game contextual effects and illusions that followed the user's movements in the game outside of the TV screen. As a result, they succeeded in stimulating the sense of movement by projecting "grids," which is one of the peripheral flow illusions, following the user's movements in the game. However, the stimuli prepared in IllumiRoom were pre-processed to match the game content, and the content could not be freely selected. In other words, it lacks versatility. Fukuchi *et al.* [9] developed "IllumiFrame". They solved the versatility problem of IllumiRoom by presenting several general illusions in the background of a music video and moving them in time with the music. However, IllumiFrame requires a reduction in the size of video content to insert illusory figures into the background. Furthermore, the subjects of their evaluation experiment feedbacked that they felt tired and sick while participating in the experiment. In order to project context-based video around the TV screen, a method that automatically generates such effects using deep neural networks and projects them outside the TV screen using a dedicated projector has also been well studied. However, such research requires the purchase and installation of a dedicated projector, placing a heavy burden on the user [15−17].

C. Zooming Operation

According to the abovementioned analysis in subsection B, stimulating peripheral vision using optical illusions usually causes the problems of low versatility and limited video size. To tackle the problems, the previous study [12] proposed zooming operations which are performed on continuous video frames to emphasize scenes with intense movement in live music videos. There were three types of zooming operations used in the study: (1) instantaneous zooming, (2) two steps zooming, and (3) gradual zooming. The evaluation results revealed that the gradual zooming improved the sense of presence. However, the zooming operation was performed in the center of the video frame, resulting in an unnatural zooming effect if no music performer was in the center of the screen. Consequently, it increased the viewers' discomfort in some videos.

III. Methodology

A. Proposed Method

For zooming operation in the previous study [12] stated in Section II, the discomfort increases when the zooming region differs from the one the viewer expects. There are two cases for this as follows:

(1) Although the intense movement of a music performer of interest was correctly detected, the zooming was performed in the center of video frames even if the music performer was not in the center of the screen.
(2) The intense movement of no interest, such as camera motion and herd movement, was detected, and the zooming was performed in the center of video frames as in item (1).

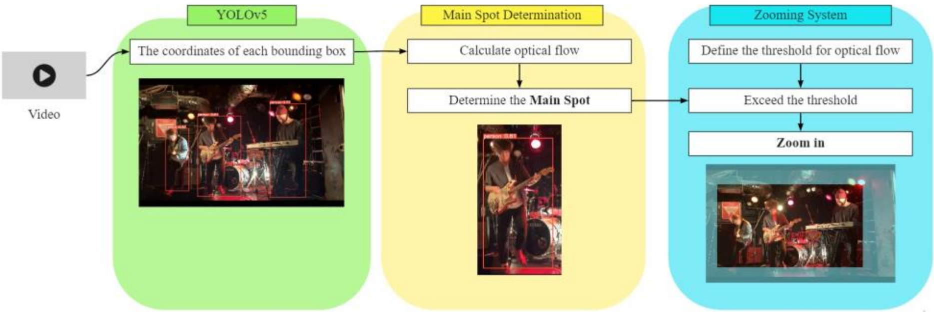


Figure 2. Overview of proposed method.

Case (2) indicates that the zooming timing was incorrect. From the above investigation, it is hypothesized that if a region of interest with intense movement, which is the "Main Spot", is correctly detected and the gradual zooming is performed at the right time, the sense of presence is improved and the discomfort of the viewers is alleviated. The reason for alleviating the discomfort is that the Main Spot is nothing but the Region of Interest (ROI) in the video that attracts the viewer's attention. To achieve these two challenges, a new zooming method (as depicted in Fig. 2) for live music videos is proposed as follows:

(1) Music performers in a video scene are detected by an object detection algorithm.
(2) Each average optical flow value identifies the most intensely moving performer among the detected music performers.
(3) The region of the video frame where the most intensely moving music performer exists is determined as the Main Spot.
(4) Based on the average optical flow value of the Main Spot compared to a certain threshold, the right time to start zooming and the right duration to perform zooming are determined.
(5) Zooming is performed on the Main Spot following the strategies determined in item (4).

The details of the above procedure are explained in the following subsections.

B. *Main Spot*

The algorithm for determining the Main Spot is shown in Fig. 3. In this algorithm, the person 0, person 1, and person 2 correspond to the left, center, and right music performers in Fig. 4, respectively. This algorithm was applied every 300 video frames to a 30FPS (Frame Per Second) video.

1) Object detection

To determine the Main Spot, its associated music performer must be detected. In this study, YOLOv5 [18] (hereafter YOLO), is used to detect music performers. YOLO is an object detection algorithm that can also detect persons in a video scene. YOLO provides bounding boxes surrounding detected music performers in continuous video frames. Although there were four music performers in the videos used in this study, only three music performers excluding the drummer were detected. This is because the drummer was too small and did not stand out due to being hidden by the vocalist. In this case, it is not necessary to focus on the drummer since the drummer is not suitable for the music performer in the Main Spot.

Algorithm1 1	:	Main spot Determination
Input	:	Video
Output	:	Main spot
1	:	**Read** bounding box coordinates for each person
2	:	**Read** Location of each person from bounding box coordinate
3	:	**While** Frame Number ≠ end
4	:	**If** Frame Number % 15 = 0
5	:	**Initiate** Optical flow calculation
6	:	**Calculate** X, Y, Z ← Average of optical flow for each person
		List person 0 ← results of X
7	:	List person 1 ← results of Y
		List person 2 ← results of **Z**
		Person0_Ave ← Average of person 0
8	:	Person1_Ave ← Average of person 1
		Person2_Ave ← Average of person 2
9	:	**Compare** Person0_Ave, Person1_Ave, Person2_Ave
10	:	Main spot ← Person_Max
11	:	**Return** Main spot

The size of the bounding box obtained from YOLO varies in each video frame. Therefore, the reference bounding box must be defined to calculate the average optical flow value in the obtained bounding box for a music performer. Otherwise, it would be impossible to track the average optical flow value of the music performer over the video frames. In this study, the reference bounding box is defined as the bounding box in which three music performers are detected for the first time. In the previous study [12], the Lucas-Kanade method [19] was used to obtain the optical flow values because it is suitable for acquiring global motion in a video, such as a camera motion and herd movement. However, it is unsuitable for capturing local movement within a bounding box. Therefore, the Gunnar-Farneback method [20] is used in this study to obtain the optical flow values.

Figure 3. Numbering of persons

2) Main spot determination

In the previous study [12], zooming was performed when the average optical flow value of the entire video screen continuously exceeded the threshold value. Therefore, the zooming was performed not focusing on music performers who were moving intensely. In this study, after identifying the Main Spot by comparing the average optical flow value of each music performer's bounding box, the timing of zooming is determined by tracking only the average optical flow value of the Main Spot over the video frames.

Additionally, the Main Spot may change several times in a video. In a 30 FPS video used in this study, the average optical flow value of each person's bounding box was observed in each video frame as shown in Fig. 5. Fig. 5 shows that the person who is moving most intensely is not always the same in the video named "video1_1".

Furthermore, the average optical flow value is highly variable. Interestingly, the person with the peak average optical flow value is replaced about every 300 frames. Therefore, it is assumed that intense movements last a maximum of 300 video frames, and the Main Spot is defined as the music performer's bounding box with the summed largest average optical flow value during the continuous 300 video frames.

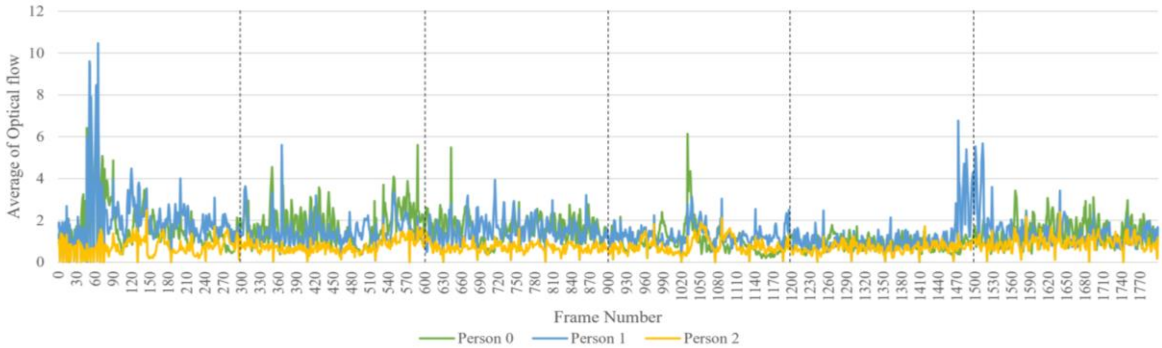


Figure 4. Variation of average optical flow value in video1_1.

In this study, the average optical flow value is used to detect a music performer's intense movement. It is assumed that the viewer also perceives the music performer's movement as intense when the average optical flow value is large. However, the average optical flow value varies depending on the video frame interval to compute it. For this reason, the observation video interval of the average optical flow value is significant when detecting the music performer's intense movement, such as jumping or stepping. Hence, a preliminary experiment was conducted to investigate the relation between the average optical flow value computed with different video frame intervals and the degree of the viewer's perception of "moving intensely" when watching the video. Three subjects (two males and one female) participated in the experiment.

The subjects were asked to respond to the video frame interval used to calculate the average optical flow value used when they felt the music performer was moving most intensely. The video frame intervals used in the experiment were 1, 7, 15, 22, and 30.

The results showed that when the average optical flow value was calculated with the video frame interval of 15, the subjects also perceived the music performer moving most intensely when the value was the highest. Therefore, in this study, the video frame interval is set to 15 when calculating the average optical flow value for each bounding box. Then, the bounding box with the largest value is chosen, as the Main Spot.

C. Zooming Method

After determining the Main Spot, zooming is performed on the Main Spot when the average optical flow value exceeds a certain threshold. In this subsection, the threshold of the average optical flow value for zooming (henceforth referred to as the "zooming threshold"), and how to perform zooming on the Main Spot are explained.

1) Zooming threshold determination

Another preliminary experiment was conducted to determine the zooming threshold of the Main Spot that determines when to perform zooming. The average optical flow value of the Main Spot in each video (12 videos in total) was analyzed in this study. The results showed that the average optical flow value of the Main Spot had a similar variation range in all videos. This may be because they were music live performances performed by the same music band members. Fig. 6 shows one of the results obtained from a video, named "video1_1". It also shows how the Main Spot varied from time to time. The horizontal axis indicates the number of video frames which corresponds to the duration time of the video, while the vertical axis indicates the average optical flow value of the Main Spot calculated with the video frame interval of 15.

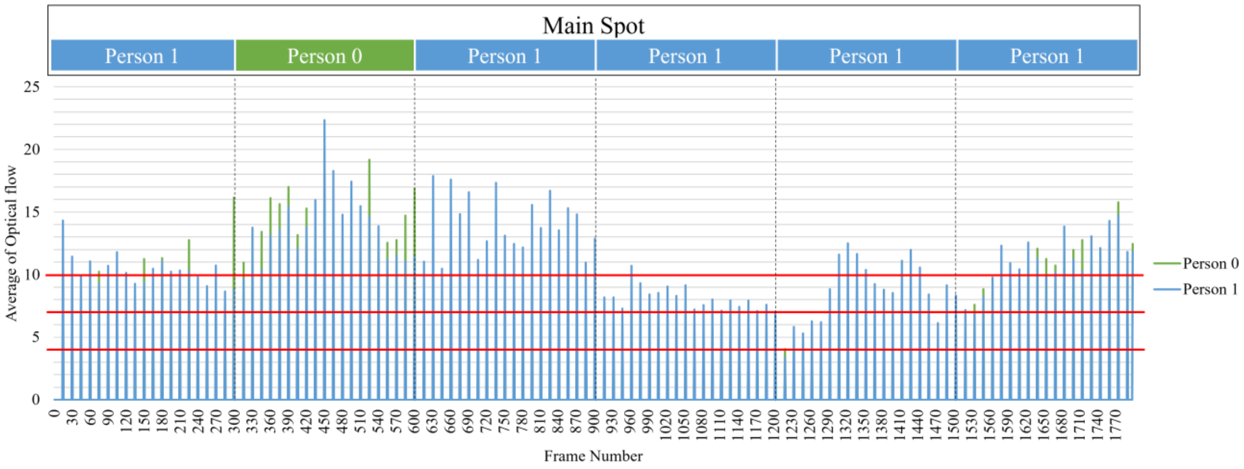


Figure 5. Average optical flow value of Main Spot in video1_1 calculated with video frame interval of 15.

If the duration of zooming is too short, it has the potential to cause discomfort to the viewer. Therefore, the duration of zooming on a professional music band's live video uploaded to YouTube was investigated. As a result, it was found that the duration of zooming was more than one second. The video frame rate used in this study is 30FPS, hence, the zooming operation needs to last more than 30 video frames. In other words, when an intense movement of the Main Spot lasts more than 30 video frames, zooming will be performed on the Main Spot. Note that zooming starts at the video frame, where the average optical flow value of the Main Spot exceeds the zooming threshold.

Fig. 6 also shows important information about the zooming threshold. If the zooming threshold is set to 4, the Main Spot (Person 0 or Person 1) is repeatedly zoomed all the time. If the zooming threshold is set to 10, zooming is not performed on the Main Spot in the frame from 900 to 1,500. Thus, the zooming thresholds of 4 and 10 are critical. The zooming threshold influences the sense of presence and the discomfort of the viewers. Therefore, the abovementioned zooming thresholds will be evaluated in the evaluation section.

2) Zooming on Main Spot

After determining the timing of zooming using the zooming threshold, the zooming is performed in the continuous video frames. In the previous study [12], zooming was performed in the center of the video screen. However, this method did not emphasize the movement of the person who should be focused on. Thus, in this study, zooming is performed on the Main Spot. In other words, zooming is performed in the center of the bounding box. The magnification rate depends on the height of the bounding box at the start of zooming, so that the height of the bounding box becomes the maximum height of the screen at the end of zooming.

Bounding boxes are obtained by YOLO every other frame and vary in size. Therefore, in this study, the height and width of the bounding box are fixed to the ones of the reference bounding box as stated previously.

IV. Evaluation

A. Experimental Setup

An evaluation experiment was conducted to verify if the proposed method improves the sense of presence and alleviates the viewers' discomfort. Twelve non-processed videos recorded at music live concerts were used in the experiment. All the videos along with descriptions, are shown in Table I.

TABLE I. Video Contents Used for Evaluation Experiment

Video1	video1_1	・Live videos taken by an amateur band ・one min, 30 FPS ・Each playing six different songs. ・Camera is handheld, but the photographer's position rarely moves.
	…	
	video1_6	
Video2	video2_1	・Live videos taken by the same amateur band as video1 ・one min, 30 FPS ・Each playing six different songs. ・Camera is handheld, but the photographer's position rarely moves.
	…	
	video2_6	

The music bands performing in the video1 and video2 series were the same, but the live venue and the performance were different. Six songs of one minute were selected from each series. There were 16 participants (12 males and four females) who were asked to watch 12 non-processed videos and 36 processed videos (three different zooming thresholds for zooming were used for the 12 non-processed videos).

After watching the videos, the participants answered questionnaires for subjective evaluation (5-point scale), concentrating on the sense of presence and discomfort. Fig. 7 illustrates the questionnaires using Google form used for the evaluation experiment. Each participant used their own laptop computer (display resolution: 1920pixel × 1080pixel) and earphones or headphones to watch videos. The experiment flow is shown in Fig. 8. Before the evaluation, a tutorial was conducted to familiarize the participants with the process of watching videos and answering questionnaires. In the tutorial, the participants watched an unprocessed short live video of 20 seconds or less and responded to the sense of presence and discomfort.

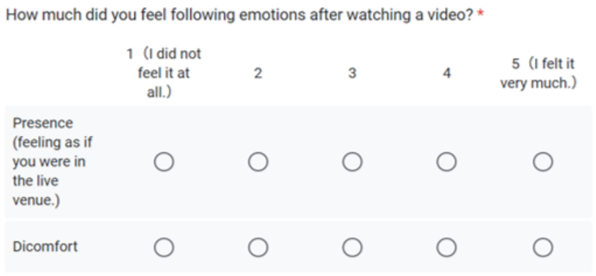


Figure 6. Questionnaires using Google form used for evaluation experiment.

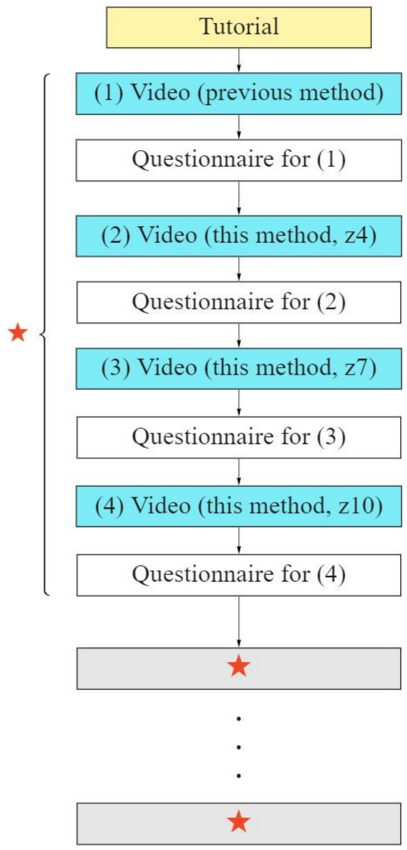


Figure 7. Flow of evaluation experiment watching videos and answering questionnaires.

In Fig. 8, "(1) Video (previous method)" means the video processed using the gradual zooming proposed in the previous study [12]. This is because the gradual zooming achieved the best performance among the three methods in the previous study. Then, "Video (this method, zN)" means a video processed using the proposed method with the zooming threshold set to N (N=4, 7, 10). The flow described as the star was repeated for six different videos in one series. The order of videos watched was the same for all the participants.

B. Evaluation Results

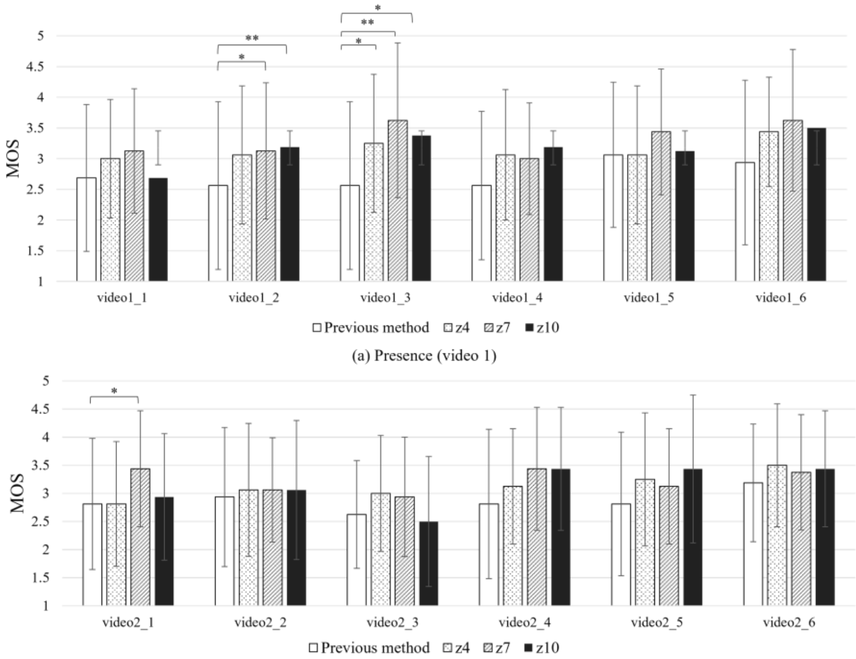
Fig. 9(a) and 9(b) depict the subjective evaluation results of the sense of presence for the video1 and video2 series, respectively. Zooming threshold values of 4, 7, and 10 are denoted by z4, z7, and z10, respectively. As mentioned in the methodology section, the zooming thresholds of 4 and 10 were selected for evaluation. In addition to them, the zooming threshold of 7 was also evaluated. In each histogram, the error bar indicates the standard deviation of the "Mean Opinion Score" (hereafter MOS).

Overall, the results show that the proposed method improved the sense of presence compared to the previous method for all videos. Particularly, the t-test results show significant differences in the six videos denoted by asterisks (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

Focusing only on the results with significant differences, the sense of presence improved the most when the zooming threshold of 7 was applied.

The discomfort results for the video1 and video2 series are shown in Fig. 10(a) and 10(b), respectively. As seen from those figures, the discomfort was alleviated by the proposed method compared to the previous method in some videos, such as video1_1, video1_3, and video1_6. However, in the rest of the videos, the previous method achieved less discomfort than the proposed method.

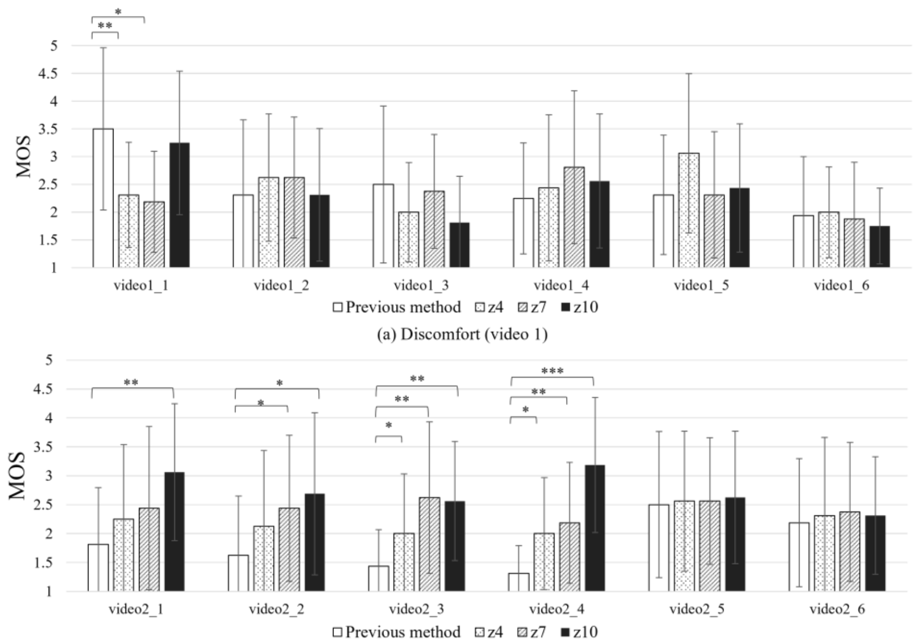
The t-test results show significant differences in the 11 videos denoted by asterisks (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). These results indicate that significant differences in the alleviation of discomfort by the proposed method were observed only in video1_1 with the zooming threshold of 4 and 7. Here, there is an important point in the results. In almost all the videos, the MOS values in the proposed method are very low such as 2.0 or 2.5, which means the viewers did not feel discomfort so much. This will be discussed in the next section.



(a) Presence (video 1)

(b) Presence (video 2)

Figure 8. Experimental results on sense of presence.



(a) Discomfort (video 1)

(b) Discomfort (video 2)

Figure 9. Experimental results on sense of discomfort.

V. Discussion

According to the result of the evaluation experiment stated in the previous section, the proposed method significantly improved the sense of presence in all the videos in the video1 and video2 series. This indicates the effectiveness of the "Main Spot" based zooming operation.

As for discomfort, focusing on the results with significant differences, the MOS value in the previous method is less than the one in the proposed method except for video1_1 with the zooming threshold of 4 and 7. It is true that the previous method is better than the proposed method. However, looking at the results of the proposed method carefully, the MOS values are very low. More than 80% of the values are less than 3.0, and more than 60% are less than 2.5 (5-point scale). The proposed method was unable to alleviate the discomfort compared to the previous method. However, it can be concluded that the sense of presence was greatly improved even though the discomfort was almost completely suppressed.

There are two possible reasons for the low MOS values for discomfort in the proposed method.

Figure 10. Music performer is out of screen when zooming

First, as shown in Fig. 10, there were some videos in which the music performer in the Main Spot protruded from the screen when zooming. This is because the coordinates of the bounding box were fixed to those in the video frame in which the three music performers were detected for the first time, as stated in the methodology section. Hence, if the music performer in the Main Spot is larger than the bounding box in later video frames, a part of the performer's body will protrude from the screen. In addition, if the music performer in the Main Spot moves to the left or right a lot or if the camera is shaken a lot, such a protrusion will also occur. Second, the duration of zooming was too short such as one second in some zooming occurrences. Table II shows the number of times a part of the music performer's body protruded from the screen and the number of times around one-second zooming occurred in the videos for which there were significant differences in the discomfort. Although increases and decreases in discomfort were also observed in other videos, they are not discussed here because there were no significant differences. Areas highlighted in yellow show videos with decreased discomfort while areas highlighted in blue show videos with increased discomfort compared to the previous method. Table II shows that, in the videos with increased discomfort, except video2_3 with the zooming threshold of 4, either a part of the music performer's body protruded from the screen, or around one-second zooming occurred, or both. In the previous method, the zooming itself did not occur in video2_3. This is because the way of obtaining and calculating the average optical flow value differed from the proposed method in this paper, and zooming was deemed unnecessary.

The zooming threshold of 7 was the best in this study to improve the sense of presence. The proposed method performs zooming when the average optical flow value exceeds the threshold value. When the zooming threshold was set to 4, the duration of zooming was too long, while the zooming threshold of 10 caused too infrequent zooming. It can conclude that the zooming threshold of 7 achieves both the appropriate duration of zooming and the appropriate frequency of zooming.

When the zooming threshold of 10 was applied, the discomfort significantly increased compared to the previous method. As shown in Table II, around one-second zooming always occurred with the zooming threshold of 10, which is one of the reasons for discomfort. When the zooming threshold is set to a large value such as 10, the average optical flow value satisfying it appears infrequently. It shortened the duration of zooming, causing around one-second zooming. It is concluded that large zooming threshold values should not be used because it would induce discomfort.

TABLE II. CHARACTERISTICS OF VIDEOS SHOWING SIGNIFICANT DIFFERENCES IN DISCOMFORT

	Zooming threshold	Protrusion from screen	Around one-second zooming
video1_1	4	0	0
	7	0	0
video2_1	10	1	2
video2_2	7	2	0
	10	1	2
video2_3	4	0	0
	7	0	2
	10	0	2
video2_4	4	2	0
	7	2	0
	10	1	3

VI. CONCLUSION

In this paper, a new zooming method, which effectively emphasizes a music performer's intense movement in music live videos by introducing the concept of "Main Spot", was proposed not only to improve the sense of presence but also to alleviate the discomfort of the viewers. The proposed method correctly identified "a music performer with intense movement", started zooming at the right time, and continued it during the right duration. The evaluation experiment results showed that the proposed method significantly improved the sense of presence compared to the previous method. As for the discomfort, the proposed method couldn't outperform the previous method. However, the result of the proposed method showed that the degree of discomfort was still very low. Hence, it can be concluded that the sense of presence was greatly improved even though the discomfort was almost completely suppressed.

Although the degree of discomfort is low, it still occurs because the music performer in the Main Spot protruded from the screen when zooming. This is because the size of the bounding box was fixed. When the photographer moved forward over recording the video, the music performer in the Main Spot was zoomed more than necessary Another discomfort factor is a frequent too short zooming (around one-second) occurrence. These two problems can be solved by redefining the reference bounding box and optimizing the zooming duration in the future.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Ai Oishi, Eiji Kamioka, Phan Xuan Tan, and Manami Kanamaru performed research; Ai Oishi Phan Xuan Tan, and Eiji Kamioka analyzed the data; Ai Oishi Phan Xuan Tan, and Eiji Kamioka wrote the paper; all authors had approved the final version.

REFERENCES

[1] W. A. IJsselsteijn, H. De Ridder, J. Freeman, and S. E. Avons, "Presence: Concept, determinants, and measurement," *Human Vision and Electronic Imaging V*, vol. 3959, pp. 520–529, 2000.

[2] M. Lombard and T. Ditton, "At the heart of it all: The concept of presence," *Journal of Computer-Mediated Communication*, vol. 3, no. 2, JCMC321, 1997.

[3] K. K. Hyong, "Maximizing presence in projection-based virtual reality systems," Phd dissertation, University of California, Santa Barbara, 2021.

[4] H. Ando *et al*., "Perceptual and cognitive mechanisms of presence and its evaluation technology," *Review of the National Institute of Information and Communications Technology*, vol. 56, no. 1.2, pp. 157–165, March 2011.

[5] Y. Kinoshita, M. Ito, K. Ozawa, and M. Morise, "Time-series analysis and model development of audience emotion for evaluation of the presence in audio-visual content," *Transactions of Japan Society of Kansei Engineering*, vol. 15, no. 3, pp. 407–414, 2016.

[6] T. Brandt, J. Dichgans, and E. Koenig, "Differential effects of central versus peripheral vision on egocentric and exocentric motion perception," *Experimental Brain Research*, vol. 16, no. 5, pp. 476–491, 1973.

[7] B. R. Jones, H. Benko, E. Ofek, and A. D. Wilson, "IllumiRoom: peripheral projected illusions for interactive experiences," in *Proc. SIGCHI conference on Human Factors in Computing Systems (CHI 2013)*, July 2013, pp. 869–878.

[8] B. R. Jones, H. Benko, E. Ofek, and A. D. Wilson, "IllumiRoom: Immersive experiences beyond the TV screen," *Communications of the ACM*, vol. 58, no. 6, pp. 93–100, 2015.

[9] T. Fukuchi, K. Matsui, and S. Nakamura, "Realization on enhancement method of content viewing experience by illusion in peripheral vision," *Information Processing Society of Japan*, vol. 59, no. 2, pp. 681–689, February 2018.

[10] T. Prokop, M. Schubert, and W. Berger, "Visual influence on human locomotion modulation to changes in optic flow," *Experimental Brain Research*, vol. 114, no. 1, pp. 63–70, 1997.

[11] W. H. Warren and K. J. Kurtz, "The role of central and peripheral vision in perceiving the direction of self-motion," *Perception & Psychophysics*, vol. 51, no. 5, pp. 443–454, 1992.

[12] A. Oishi, E. Kamioka, P. X. Tan, and M. Kanamaru, "Improving presence experience in live music videos by using adaptive zooming," J*ournal of Image and Graphics (United Kingdom)*, vol. 10, no. 2, pp. 70–75, 2022.

[13] T. Miura, "Visual Attention and traffic safety focusing on useful field of view," *Journal of the Illuminating Engineering Institute of Japan*, vol. 82, no. 3, pp. 180–184, 1998.

[14] Y. Okano, Y. Saiga, Y. Hashimoto, T. Nojima, and H. Kajimoto, "A study on the display method to the peripheral view for augmentation of speed sensation," *IPSJ SIG Technical Reports*, pp. 145–150, 2008.

[15] N. Kimura and J. Rekimoto, "ExtVision: Augmentation of visual experiences with generation of context images for a peripheral vision using deep neural network," in *Proc. 2018 CHI Conference on Human Factors in Computing Systems*, April 2018, pp. 1–10.

[16] M. A. Shirazi, H. Cho, and W. Woo, "Augmentation of visual experiences using deep learning for peripheral vision based extended display application," in *Proc. 2019 IEEE International Conference on Consumer Electronics (ICCE)*, IEEE, January 2019, pp. 1–2.

[17] M. A. Shirazi, R. Uddin, and M. Y. Kim, "Supervised learning based peripheral vision system for immersive visual experiences for extended display," *Applied Sciences*, vol. 11, no. 11, 4726, 2021.

[18] Yolov5. [Online]. Available: https://github.com/ultralytics/yolov5

[19] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in P*roc. 7th IJCAL*, 1981, pp. 674–679.

[20] G. Farneback, "Two-frame motion estimation based on polynomial expansion," in *Proc. Scandinavian Conference on Image Analysis*, 2003, pp. 363–370.