Improving Presence Experience in Live Music Videos by Using Adaptive Zooming

Ai Oishi, Eiji Kamioka, Phan Xuan Tan, and Manami Kanamaru
Graduate School of Engineering and Science, Shibaura Institute of Technology, Tokyo, Japan
Email: {ma21024, kamioka, tanpx, kanamaru.manami.a4}@shibaura-it.ac.jp

***Abstract*—The demand for live music video has been increasingly reported. To bring high satisfaction to the users, it is necessary to improve the presence experience while watching the videos. Our previous study introduced a method which suddenly performs a zooming operation to the video frame where a high intensity of motion is detected, resulting in better presence. However, such a sudden zooming operation makes the users feel uncomfortable, leading to low score of pleasure. In this paper, two types of adaptive zooming operations are proposed not only to enhance the presence experience but also to eliminate such an uncomfortable feeling. One adapts zooming magnitude to the intensity of motion. The other adapts zooming speed to human's visual perception, that is to say, the zooming is performed gradually. The experimental result shows that the proposed operations outperform the previous simple zooming operation in terms of the presence experience and comfortability.**

***Index Terms*—presence, activity, peripheral vision, live music video**

I. Introduction

In an interview in 2002, David Bowie [1] stated the growing importance of streaming services and live music. Now it becomes true: Spotify, YouTube and other streaming services turned music into running water while simultaneously the growth of the live music industry accelerated. The live music itself as an artistic uniqueness is a driving motivation for visiting concerts because of many reasons, for examples, artist worship, the need for live performances, the value of being close to artist, the urge for just being there and so on [2]. However, people do not always have a chance to directly attend such events. Instead, they rely on live streaming services even though the experience is probably different from the real one. To bring high satisfaction to live streaming users, it is necessary to improve so-called feeling of "being there", in other words, presence experience [3], [4]. According to Ando *et al.* [5], presence is composed of multiple sensory components: Spatial component, Temporal component and Physical component. "Activity", which is one of the factors of temporal components, indicates a sense of change and a sense of motion, and has a correlation with presence [6]. Therefore, some existing studies [7], [8] attempted to improve the presence experience in gaming by stimulating peripheral vision to capture the sense of motion. This is because the peripheral vision is sensitive to motion. Indeed, optical flow in the peripheral vision is considered to play an important role in the perception of motion [9]. In those studies, optical illusion was leveraged to activate the peripheral vision. However, there is a high possibility that such an illusion makes the users feel uncomfortable. To deal with this problem, alternatively, our previous study [10] introduced zooming operation on specific video frames to activate peripheral vision, resulting in higher presence experience. Nevertheless, such a sudden video frame expansion still made the users uncomfortable.

In this paper, two types of adaptive zooming operations are proposed not only to enhance the presence experience but also to eliminate such an uncomfortable feeling mentioned above. One performs a zooming to the video frames, adapting the zooming rate to the intensity of motion. The other performs a zooming gradually to successive video frames where the intense motion is detected. The experimental result revealed that the proposed operations enhanced the presence experience and eliminated uncomfortable feeling more than our previous study [10].

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

II. Background

A. Presence

Presence – a psychological effect is the sense of "being there" in a mediated environment [3], [4]. The presence can be affected by "realism", "transportation" and "immersion" [4].

The presence can also be defined as a multidimensional sense composed of various elements (see Fig. 1). According to the study by Ando *et al.* [5], the sense of presence is formed by three components, which are Spatial component, Temporal component and Physical component. In order to enhance the sense of presence in live music videos, it is necessary to clarify which elements of which components should be enhanced. In an existing study, Kinoshita *et al.* [6] showed a strong correlation between "Deactivation - Activation" and the presence experience. Therefore, in this paper, "Activity", which is an element of Temporal component in Fig. 1, is focused on. "Activity"

Manuscript received October 15, 2021; revised April 2, 2022.

doi: 10.18178/joig.10.2.70-75

can be interpreted as the sense of change or the sense of motion. This is a sensitivity with which you feel like moving or you feel dynamism when you watch videos.

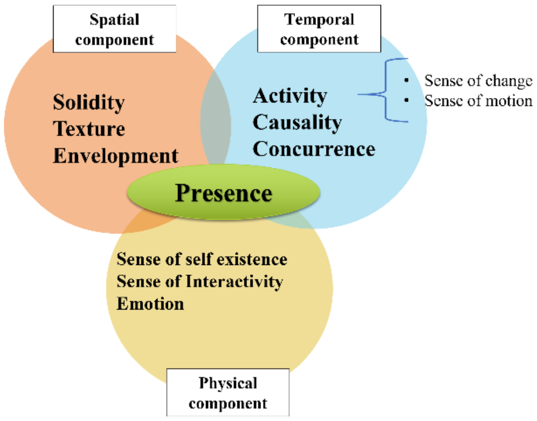

Figure 1. Components of presence [5].

B. *Peripheral Vision*

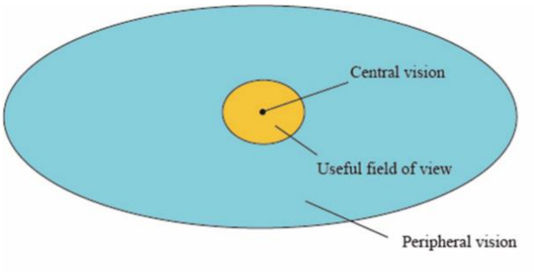

Figure 2. Schematic diagram of human vision.

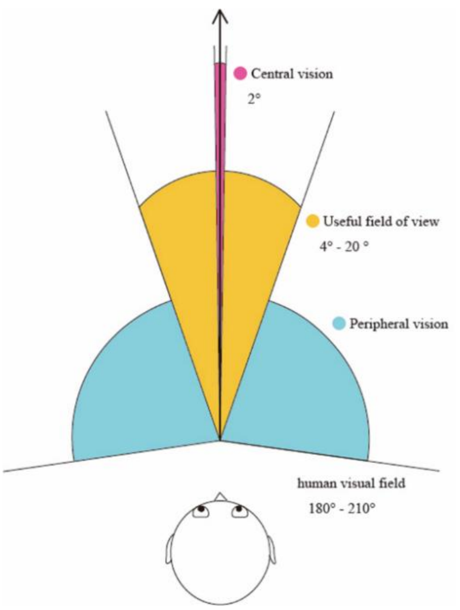

Figure 3. Range of human vision.

According to Miura [11], the human visual field spans 180 to 210 degrees. The central field of vision, which has high resolution, is only about 2 degrees. The area around the central vision where human can perceive objects clearly is called the useful field of view. This range varies depending on psychological factors. The field of view other than the central vision including the useful field of view is called the peripheral vision (see Fig. 2). The range of visual field can be described in Fig. 3. In addition, the central and peripheral visions play different roles [9]. It is what they perceive. The central vision is known as sensitive to color and insensitive to movement. On the other hand, the peripheral vision is better able to perceive the motion with high-velocity stimuli than the central vision. In other words, the peripheral vision is insensitive to color and brightness and sensitive to motion [12]. Therefore, optical flow in peripheral vision makes human perceive the motion. Using this theory, some studies revealed that peripheral vision can be activated by stimuli such as optical illusions [7], [8], [12], [13].

III. Improving Presence Experience by Activating Peripheral Vision

A. *Optical Illusion*

Optical illusion is a kind of illusion and illusory figures using this illusion give human a special perception. Several studies have reported that the optical illusion around a display monitor can stimulate human peripheral vision. Such an activated peripheral vision makes human feel motion. Jones *et al.* [7] used the optical illusion to develop “IllumiRoom”. They provided not only illusions but also special effects outside a television monitor in order to activate the peripheral vision for improving the sense of motion. According to their study, “Grid”, which is one of the peripheral flow illusions, seems to be successful at impacting a sense of apparent motion. Note that “Grid” is composed of infinite grids and moves with the motion in the game. Such illusions have made users feel sick. Furthermore, this system has a problem in terms of general versatility. These illusions have already been processed, thus, “IllumiRoom” cannot be used for freely selected content. Fukuchi *et al.* [8] developed “IllumiFrame” to solve this problem by using optical illusions. They have built a system that can be applied to a variety of content, e.g. music videos, by automatically presenting the illusion to music. However, this system makes images unnatural because it is a smaller version of the original image with optical illusions inserted in the background. Some users told that it made them tired and sick. Therefore, a new system, which does not make users uncomfortable with highly keeping the sense of motion, is needed. In our previous study [10], a zooming operation was proposed to stimulate human peripheral vision. When a zooming is performed in the watching video, the images are suddenly enlarged, resulting in changing in peripheral of images. “Changing in peripheral of images by zooming can activate peripheral vision” was hypothesized in that study. It was verified by subjective evaluation. As the result, the presence experience was improved. However, the sense of “Pleasure” of the users also decreased. “Pleasure” is one of the emotion, which is an element of Physical component in presence. According to Kinoshita *et al.* [6], there is a weak negative correlation between the presence experience and “Unpleasant - Pleasant”. Therefore, the zooming operation must be improved to keep a certain level of the sense of “Pleasure”.

B. *Hypothesis*

The conclusions of our previous study [10] revealed the followings: First, zooming operation can activate the peripheral vision, then improve the sense of motion and the presence experience. However, the sense of “Pleasure” decreases. Second, the heightened sense of “Displeasure” does not always affect all the components of presence. Third, when a zooming operation is performed frequently, the higher sense of “Displeasure” is observed. To achieve the goal of this study, a hypothesis can be made as follows: If zooming operation to video frames can be performed more naturally, it would be possible to achieve the high presence experience keeping the sense of “Pleasure”. Therefore, in this paper, two types of adaptive zooming operations are proposed to perform zooming operation more naturally.

IV. METHODOLOGY

According to the report by Fukue *et al.* [14], even though the same recording and playback systems are used, there are differences in the sense of presence felt by the users depending on the audiovisual content. Therefore, it is necessary to examine the elements of sensibility evoked from the video content and identify which elements are insufficient for the sense of presence. Thus, a preliminary experiment was conducted to investigate such insufficient elements of sensibility in conventional live music videos [10]. If those insufficient elements of sensibility are enhanced in some way, the sense of presence could be improved. The result of the preliminary experiment showed that the sense of "Activity" was insufficient in live music videos among the components of the sense of presence shown in Fig. 1.

To enhance the sense of “Activity” in this study, optical flow in peripheral vision have been used [10] as several existing studies [7], [12], [13] did. Specifically, when an intense movement is detected, the video frame is enlarged, resulting in giving a strong stimulation to the peripheral vision by zooming.

In addition to our previous study [10], in this paper, two types of adaptive zooming operations are proposed to improve the sense of “Pleasure” as follows:

(i) Adaptation of zooming magnitude to intensity of motion

When a video frame is enlarged by zooming, the rate of magnification is adapted to the intensity of the motion, which can be detected by the image difference between the adjacent video frames (See Pattern B in Table I). This is intuitively understandable because humans perceive nearby moving objects larger and faster than faraway moving objects. Therefore, this zooming operation must be natural to the users.

(ii) Adaptation of zooming speed to human’s visual perception

When a video frame is enlarged by zooming, the zooming speed is adapted to human’s visual perception. More concretely, the enlargement is gradually performed to successive video frames (See Pattern C in Table I). If a video frame is suddenly enlarged from the original one, it would be unnatural. When the enlargement is performed by taking time, this zooming operation must be natural to the users.

The above two adaptive zooming operations are compared with the simple zooming operation in our previous study (See Pattern A in Table I) [10].

The operation of zooming is triggered based on the threshold of optical flow value which is estimated by OpenCV library [15]. Then, when an optical flow value larger than the threshold is detected, the video is enlarged by zooming. More concretely, the following two algorithms described by pseudo-code are used for this purpose.

Frame Determination Algorithm:

Input : Video frames
Output : F_z: Frames to be zoomed in
Define X: the magnitude of optical flow vector
Define M: the rate of X (=90%)
  Initiate Optical Flow calculation
    Determine V**:** the rate of moving video frames
      Read Video frames (Until end of the video)
      If V > M
        tmp_zoom = 1
      else tmp_zoom = 0
      Initiate F_z (List of tmp_zoom)
  Return F_z

Zooming Algorithm:

Input : F_z
Video frames
Output : Processed video
Define T: threshold for zooming
Call F_z by Frame determination function
Initiate Continuous judgment processing
  Fix F_z
  Initiate F_z_continuous
Read Video frames (Until end of the video)
  If F_z_continuous [i] = 1
    Resize the frame using T
    Write the resized image

TABLE I. ZOOMING PATTERN

Label	Name	Description	Reasons
Pattern A	Simply zooming	Simply zooming video frame when intense motion is detected.	This is the method used in the previous study [10]. This is used to compare with other zooming methods for evaluation.
Pattern B	2-stage zooming	Slightly zooming video frame when intense motion which is smaller than pattern A is detected. Simply zooming video frame when intense motion which is the same level as pattern A is detected.	Using two different methods for zooming in order to adjust zooming magnitude to the level of intense motion.
Pattern C	Smooth zooming	Not zooming video frame quickly like patterns A or pattern B. Instead, zooming video frame gradually when intense motion which is the same level as pattern A.	Zooming slowly to prevent users from being surprised by sudden zooming like pattern A.

In Frame Determination Algorithm, the optical flow value in each video frame is calculated. When a video frame is regarded as having intense motions, the magnitude of optical flow vector "X" is set to 1 based on the preliminary study [10]. When the number of optical flow vectors indicating intense motions (X=1) accounts for more than 90% in the video frame, "tmp_zoom" is set to 1. In other frames, "tmp_zoom" is set to 0. After determining the tmp_zoom for all the video frames, the sequence of these values composed of "0" and "1" is stored in F_z.

In Zoom Algorithm, the information of F_z is used to determine if the considered video frame is enlarged or not. As mentioned above, when the value of a component is "1", the corresponding video frame is enlarged and when the value is "0", the corresponding video frame is not enlarged. Therefore, when the different value is frequently appears in the sequence, the video image would be unnaturally zoomed in and zoomed out repeatedly. Hence, a process in which the continuity of video frames to be enlarged is optimized is introduced. Specifically, in this study, the video frames are enlarged when the value of "1" continuously appears more than 30 times. (Note that this is to keep the enlarged video frames for more than 1 second because the FPS of the used videos was about 30.) Therefore, if the value of "0" appears in a sequence of "1"s, the value of "0" is replaced with "1". On the other hand, if the value of "1" appears in a sequence of "0"s, the value of "1" is replaced with "0". Through this process, F_z is altered to an array of consecutive "1"s and consecutive "0"s, which is called F_z_continuous. In addition, the threshold for zooming, "T", is set for each zooming pattern as described in Table II. Zooming is performed by matching the size of cropped image using equations (1) and (2) to the width and height values of the original video, as shown in Fig. 4. The magnification factor for Pattern A is 1.49. This algorithm is used in Pattern A and used as the template for Pattern B and Pattern C. In Pattern B, the video frames are enlarged at a magnification factor of 1.13 even when the number of optical flow vectors with $X = 0.5$ accounts for more than 90%. In Pattern C, the video frames are gradually enlarged at a magnification factor of 1 to 1.49 while the value of "1" is continuously set in F_z_continuous. The magnification ratio for each zooming pattern is shown in Table II.

TABLE II. THRESHOLD AND MAGNIFICATION RATIO

	Pattern A	Pattern B	Pattern C
T: threshold for zooming	0.09	0.03 / 0.09	0.09
Magnification ratio	1.49	1.13/1.49	0~1.49

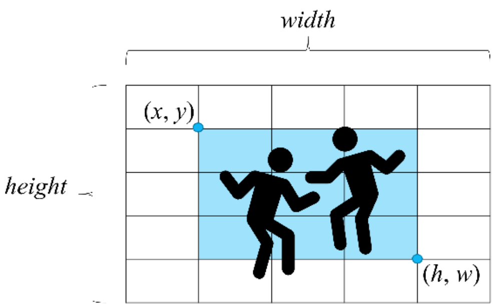

Figure 4. Magnification method.

$$(x, y)=(width * \mathrm{T}, height * \mathrm{T}) \quad (1)$$

$$(h, w)=\big(width * (1\text{-T}), height * (1\text{-T})\big) \quad (2)$$

V. EVALUATION

A. Experiment Setup

To evaluate the proposed method, an experiment was conducted. In this experiment, in total 11 videos of live music performance were prepared. These videos were then processed under Pattern A, B and C. As examples, screenshots of three videos used in the experiment are shown in Fig. 5. Fifteen participants (11 males and 4 females) were asked to watch the 11 original videos and 33 processed videos (3 zooming pattern for the 11 original videos), and also asked to answer questionnaires to provide 5-point scores for subjectively assessing their presence experience and the sense of "Pleasure". The participants used their own laptop computers (with the display resolution of 1920 pixel × 1080 pixel) and their own earphones to watch videos.

Figure 5. Videos used for evaluation.

B. Result

The scores of the presence experience and the sense of "Pleasure" are illustrated in Fig. 6 and Fig. 7, respectively. Fig. 6 shows that in most of videos, the evoked presence experience in Pattern B and Pattern C were higher than Pattern A. Even though adaptive zooming operations (using Pattern B and Pattern C) outperform our previous study (using pattern A), there are noticeable gaps between the presence scores obtained by Pattern B and Pattern C. Particularly, Pattern B provides lower presence scores in most of videos. This trend is obvious in video 1, 2 and 5 whose physical movements of the performers are extremely strong. Interestingly, in these videos, the gaps between Pattern A's scores and Pattern C's scores are also considerably large. In terms of the sense of "Pleasure" (see Fig. 7), the adaptive zooming operations (using Pattern B and Pattern C) also outperform our previous study (using Pattern A). Similar to the presence experience, there are gaps between the sense of "Pleasure" scores obtained by Pattern B and Pattern C. Particularly, Pattern C shows better result than Pattern B in most videos. In comparison with Pattern A, Pattern C provides significantly higher sense of "Pleasure" in in video 1, 2, 5 and 7 which comprise of stronger movements of the performers. However, this trend is not always kept across all the videos. In some videos, for example, video 3, 10 and 11, the opposite trend happens.

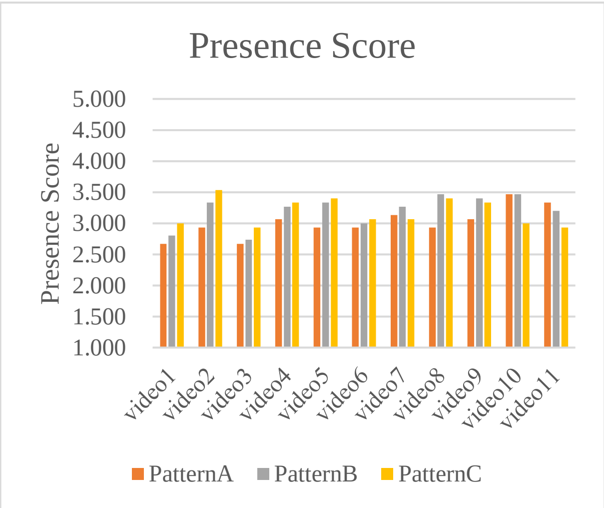

Figure 6. Presence score in each zooming pattern.

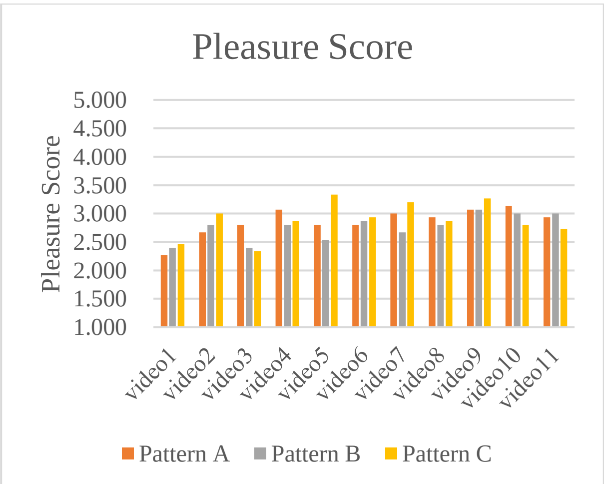

Figure 7. Pleasure score in each zooming pattern.

VI. Discussion

By using the proposed adaptive zooming operations, the presence experience has been enhanced across most of the videos, especially when Pattern C with gradual zooming was applied to original videos.

For the sense of "Pleasure", the adaptive zooming operations did not always provide consistently better results. According to Fig. 7, except video 3, 10 and 11, Pattern C achieved better sense of "Pleasure" score than Pattern B and Pattern A. It means that basically a sudden zooming operation makes the users uncomfortable. Interestingly, gradual zooming (Pattern C) provides a good result in the videos where the performers have strong movements (e.g., jump, quickly move to the left or right). Thus, the results may be related to the intensity of motion in the original live video [16]. To clarify this, the results obtained from Pattern A and Pattern C were carefully investigated. Table III tabulates the numerical result of the sense of "Pleasure" scores when Pattern A and Pattern C were applied to each original video, with the average value of optical flow. In Table III, the cells colored in pink indicate the cases where the sense of "Pleasure" score in Pattern C is higher than the one in Pattern A. Accordingly, when the optical flow value is large, for example, in video 1, 2, 5, 6 and 7, the gradual zooming operation always provides higher sense of "Pleasure" to the users. Meanwhile, when the performers do not have strong movements, the sense of "Pleasure" score in Pattern C is relatively smaller than the one in Pattern A. This can be verified in the blue cells in Table III.

TABLE III. Pleasure Score and Average of Optical Flow Vector in Original Video

Label	Average of optical flow vector	Pattern A	Pattern C
video1	0.690895385	2.267	2.467
video2	0.522153496	2.667	3.000
video3	0.470627528	2.800	2.333
video4	0.689678496	3.067	2.867
video5	0.754686886	2.800	3.333
video6	0.710671681	2.800	2.933
video7	1.482496074	3.000	3.200
video8	0.52355658	2.933	2.867
video9	0.405581201	3.067	3.267
video10	0.497891129	3.133	2.800
video11	0.623753746	2.933	2.733

VII. Conclusion

In this paper, two types of adaptive zooming operations were proposed not only to enhance the presence experience but also to eliminate uncomfortable feeling. One performs a zooming to the video frames, adapting the zooming rate to the intensity of motion. The other performs a zooming gradually to successive video frames where the intense motion is detected. Through the evaluation experiment, it was revealed that the proposed two adaptive zooming operations outperformed our previous study. In addition, it is found that there is a strong connection between the sense of "Pleasure" and the intensity of the motion in the original video. Particularly, when a sudden zooming is applied to the video with strong movements of the performers, the users feel uncomfortable with a higher possibility. On the other hand, when the proposed gradual zooming is performed, higher sense of "Pleasure" can be achieved.

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] J. Pareles, "David Bowie. 21st-century entrepreneur," *New York Times*, 2002.
[2] M. Mulder and E. Hitters, "Visiting pop concerts and festivals: Measuring the value of an integrated live music motivation scale," *Cultural Trends*, vol. 30, no. 4, 2021.
[3] W. A. IJsselsteijn, H. D. Ridder, J. Freeman, and S. E. Avons, "Presence: Concept, determinants, and measurement," *Human Vision and Electronic Imaging V*, vol. 3959, pp. 520-529, June 2000.
[4] M. Lombard and T. Ditton, "At the heart of it all: The concept of presence," *Journal of Computer-Mediated Communication*, vol. 3, no. 2, 1997.
[5] H. Ando, *et al.*, "Perceptual and cognitive mechanisms of presence and its evaluation technology," *Journal of the National Institute of Information and Communications Technology*, vol. 56, no. 1-2, March 2011.
[6] Y. Kinoshita, M. Ito, K. Ozawa, and M. Morise, "Time-Series analysis and model development of audience emotion for evaluation of the sense of presence in audio-visual content," *Transactions of Japan Society of Kansei Engineering*, vol. 15, no. 3, pp. 407-414, 2016.
[7] R. Jones, H. Benko, E. Ofek, and A. D. Wilson, "IllumiRoom: Peripheral projected illusions for interactive experiences," in *Proc. the SIGCHI Conference on Human Factors in Computing Systems*, July 2013, pp. 869-878.
[8] 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.
[9] J. J. Gibson, *The Perception of the Visual World*, Boston: Houghton Mifflin, 1950.
[10] A. Oishi and E. Kamioka., "Improvement of live music video presence using zooming," *IEICE Technical Report*, vol. PRMU2021-2, pp. 7-12, 2021.
[11] 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.
[12] 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 (HCI)*, vol. 2008, pp. 145-150, 2008.
[13] K. Nakajima and K. Fukuchi, "A velocity enhancement system for sports video based on dynamic perceptual characteristics of peripheral vision," *IPSJ SIG Technical Reports (HCI)*, vol. 2013, no. 8, pp. 1-7, 2013.
[14] K. Fukue, K. Ozawa, and Y. Kinoshita, "Study on multidimensional structure of the sense of presence in audio-visual content," *Transactions of Japan Society of Kansei Engineering*, vol. 11, no. 2, pp. 183-192, 2012.
[15] OpenCV. [Online]. Available: https://opencv.org/
[16] M. Tanaka, "Relationship between impressions of movement and temporal and spatial information," *Bulletin of the Faculty of Education*, vol. 67, pp. 39-44, 2020.

Ai Oishi was born in 1998. She received his B.E. degree from Shibaura Institute of Technology in 2021. Now she is pursuing her M.E. in Graduate School of Engineering and Science, Shibaura Institute of Technology. Her research interests include image and video processing.

Eiji Kamioka received the B.S., M.S., and D.S. degrees in physics from Aoyama Gakuin University. He is a Professor with the Shibaura Institute of Technology (SIT). Before joining the SIT, he was working with the SHARP Communication Laboratory, Institute of Space and Astronautical Science (ISAS) as a JPSP Research Fellow, and the National Institute of Informatics (NII) as an Assistant Professor. His current research interests encompass mobile multimedia communications and ubiquitous computing.

Phan Xuan Tan received B.E degree (2009) in Electrical-Electronic Engineering from Military Technical Academy, M.S degree (2013) in Computer and Communication Engineering from Hanoi University of Science and Technology (HUST) and Ph.D. degree (2018) in Functional Control Systems from Shibaura Institute of Technology (SIT). Since October 2018, he has been an Assistant Professor with SIT. His current research interests include image and video processing, 3DoF – 360 VR video & volumetric video, light field imaging, deep learning for visual computing, multimedia Quality of Experience, multimedia networking.

Manami Kanamaru is post-Doctoral researcher of Electrical Engineering and Computer Science at Shibaura Institute of Technology, her B.E (2016), M.E (2018) and D.E (2021) degrees in Engineering from Shibaura Institute of Technology.
Her current research interests include medical engineering and electric field simulation.