Human Factors in Communication of Design, Vol. 177, 2025, 58–68
https://doi.org/10.54941/ahfe1006331

Improving Watching Comfort of Live Free Viewpoint Video by Catch-Up Playback

Taisei Kudo[1], Chun Xie[2], and Itaru Kitahara[2]

[1]College of Engineering Systems School of Science and Engineering, University of Tsukuba, 1-1-1 Tennoudai, Tsukuba, Ibaraki, 305-8573, Japan

[2]Center for Computational Sciences, University of Tsukuba, 1-1-1 Tennoudai, Tsukuba, Ibaraki, 305-8577, Japan

ABSTRACT

This paper addresses the time gap issue that arises during live free-viewpoint video (FVV) viewing when video-play is paused to observe key moments from various angles. While live FVV technology allows users to switch viewpoints in real time, pausing video-play to observe key moments generates delay from real-world events, damages the feeling of sharing and unity in real time. To tackle with this problem, this paper proposes catch-up playback method to recover delays while maintaining watching comfort. Three types of catch-up playback method are introduced: Skip Playback, which instantly skips to the live point; Constant-Speed Playback, which plays back at a constant high speed until catching up; and Accelerated-Decelerated Playback, which gradually accelerates and then decelerates to normal speed. These methods dynamically adjust video-play speed based on the delay time. User experiment comparing these methods revealed that Constant-Speed Playback provided the most comfortable viewing experience, while Skip Playback received the lowest ratings. Constant-Speed Playback also outperformed Accelerated-Decelerated Playback in terms of “Easy to watch”. These results suggest that maintaining temporal continuity in video-play is crucial for ensuring a comfortable viewing experience, while continuous speed changes during catch-up playback may negatively affect watching comfort.

Keywords: Free viewpoint video, Watching comfort, Video-play speed adjustment

INTRODUCTION

The widespread adoption of streaming services has led to a growing demand for live video viewing, particularly for sports broadcasts (PwC, 2024). Among these, the utilization of free-viewpoint video (FVV), which allows viewers to freely select their favourite viewpoints, has expanded and gained significant attention. This technology enables a more active and personalized viewing experience compared to conventional viewing style with fixed viewpoints. Additionally, the “real-time-ness” of live video, where the video content is presented without or micro-time delays relative to real-world events, is a critical feature for live viewing. Live FVV (Perez et al., 2022) combines real-time video streaming with viewpoint freedom, making it particularly promising for applications in sports, such as detailed observation of baseball swings or soccer shots (Amar et al., 2023).

Two primary methods are used for viewing live FVV. The first method, as illustrated in Figure 1(a), enables viewpoint switching during live video-play without pausing the video (Koyama et al., 2003). While this approach preserves real-time-ness, it makes detailed observation of key moments from various angles difficult as the video continues to progress during viewpoint changes. As shown in Figure 1(b), the second one enables users to pause the video and freely switch viewpoints (Nagai et al., 2018). While this method facilitates detailed observation, it introduces delays in video-play due to the pause, resulting in a loss of real-time-ness. Consequently, the shared experience of simultaneous event observation, which contributes to the excitement, unity, and emotional impact of live viewing (Hertzog et al., 2020), is compromised.

This research aims to address the aforementioned challenges associated with live FVV viewing by proposing a novel playback method—catch-up playback—that mitigates delays caused by multi-angle observation while maintaining watching comfort. Here, "watching comfort" is defined as the subjective ease and enjoyment of video watching without stress or discomfort. As illustrated in Figure 1, catch-up playback dynamically adjusts the video-play speed after a pause, allowing the viewer to recover delays caused by multi-angle observation and catch up with real-world events. The effectiveness of the proposed method is evaluated through a user survey assessing the impressions of videos processed with catch-up playback.

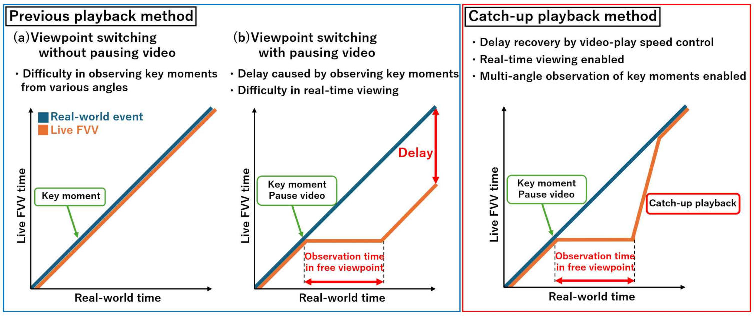

Figure 1: Catch-up playback for delay recovery. (a) When switching viewpoints without pausing the video, it becomes difficult to observe key moments in detail from multiple angles. (b) When the video is paused to switch viewpoints, video-play delay accumulates proportionally to the observation time. Once a delay occurs, resuming real-time viewing becomes challenging. Catch-up playback methods adjust video-play speed to recover delays, enabling real-time viewing even after detailed observation.

RELATED WORKS

Applications of Free-Viewpoint Video

Free-viewpoint video (FVV) technology integrates video data captured by multiple cameras to reconstruct 3D models of the target scene, which enables the generation and presentation of video from arbitrary viewpoints (Smolic et al., 2004). FVV has been applied to various fields, including sports, entertainment and education.

Sankoh et al. (2015) proposed an algorithm that delivers an immersive FVV experience, allowing viewers to observe sports scenes from arbitrary viewpoints using a single moving camera. Their approach improves FVV quality by accurately estimating homography matrices and precisely segmenting multiple target regions. Kitaguchi et al. (2024) applied FVV to surgical video recording systems, enabling clear observations of surgeries from unobstructed viewpoints by positioning multiple cameras around the surgical field.

Delay in Live Free-Viewpoint Video

Maintaining real-time-ness is critical for live FVV, leading to studies addressing delay issues. Delays caused by video generation and data transmission have been investigated, with efforts to enable low-latency live FVV (Nonaka et al., 2018).

Another type of delay arises during scene observation when switching viewpoints. Two primary interfaces for live FVV have been explored: one that allows viewpoint switching without pausing the video and another that enables free viewpoint observation after pausing the video.

Koyama et al. (2003) proposed a 3D video display system that reconstructs 3D models from videos captured by multiple cameras and transmits them to the distant place in real time. By representing subjects with lightweight 3D models composed of a single plane and 2D textures, it is possible to reduce data size, enabling real-time 3D video display. While their system supports real-time viewpoint switching, detailed multi-directional observation of subjects remains difficult.

Nagai et al. (2018) proposed an on-site visual feedback system that generates and displays FVV in real-time using the technique of bullet-time videos. Bullet-time videos sequentially switch multi-view videos along the arrangement of shooting cameras, which can create the impression of viewpoint movement. They also developed a delayed feedback system that allows detailed multi-directional observation of subjects by introducing intentional video-play delay. However, this delay undermines real-time-ness, diminishing the immersive experience.

While previous studies have focused on implementing live FVV systems, less attention has been paid to methods for recovering delays that occur during live FVV viewing.

Effects of Video-Play Speed on Viewers

The impact of video-play speed on viewers has primarily been studied in the context of educational videos. Mo et al. (2022) experimentally examined how video-play speed affects learning outcomes and cognitive load in online learning environments. Students viewed videos at different video-play speeds (1.0x, 1.25x, 1.5x, and 2.0x), with learning outcomes and cognitive load measured. Their results showed that vide-play speeds of 1.25x and 1.5x achieved the highest learning outcomes, while speeds of 2.0x significantly reduced learning efficacy. They also observed an increase in cognitive load with higher video-play speeds.

While these studies focus on cognitive effects of video-play speed, such as learning outcomes and cognitive load, they primarily address educational videos. Limited attention has been given to the impact of video-play speed on videos intended for entertainment or information retrieval. This study investigates the effects of video-play speed on watching comfort in the context of live free-viewpoint video (FVV), providing insights distinct from cognitive factors like learning outcomes or cognitive load.

CATCH-UP PLAYBACK TIME CONTROL FOR LIVE FREE-VIEWPOINT VIDEO

The proposed catch-up playback time control method for live free-viewpoint video is illustrated in Figure 2. During video-play, the video can be paused at key moments to allow detailed observation from various angles. After completing the observation, video-play resumes, and delays are recovered through catch-up playback. Three methods for catch-up playback are considered: Skip Playback, Constant-Speed Playback, and Accelerated-Decelerated Playback. These methods are categorized into discontinuous playback and continuous playback based on time continuity.

1. Discontinuous Playback (Skip Playback)

Skip Playback instantly skips to the live point (the latest point in the stream) upon resuming the video. While this method eliminates delays by skipping the corresponding video segments, it introduces temporal discontinuity in the video-play timeline.

2. Continuous Playback

2-1 Constant-Speed Playback

Constant-Speed Playback recovers delays by resuming the video at a fixed high-speed playback rate until reaching the live point. This method preserves temporal continuity in the video timeline. However, the video-play speed changes discretely, meaning the speed transitions are not continuous.

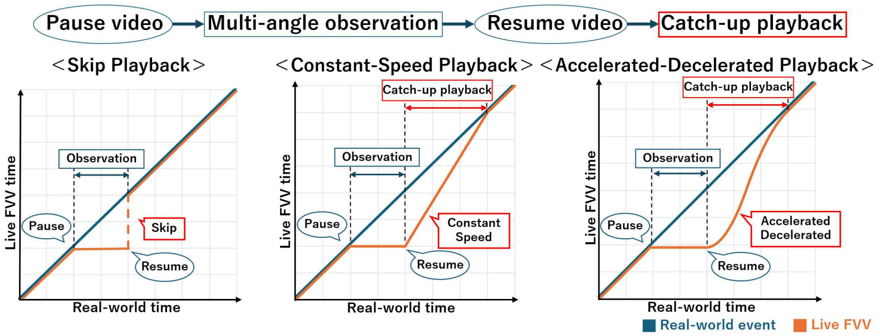

Figure 2: Delay caused by observing key moments is recovered by catch-up playback after video resumption. Three types of catch-up playback methods: skip playback, constant-speed playback, and accelerated-decelerated playback are illustrated.

2-2 Accelerated-Decelerated Playback

Accelerated-Decelerated Playback gradually increases the video-play speed upon resumption to approach the live point and then decelerates back to normal speed (1x). This method maintains both temporal continuity in the video timeline and continuous transitions in video-play speed.

CATCH-UP PLAYBACK PROCESSING

Skip Playback

Figure 3 illustrates the changes in video-play time during Skip Playback. In this method, delays caused by observing key moments are recovered by skipping to the live point immediately after resuming video-play. During observation, the video is paused to allow multi-directional observation of key moments by switching viewpoints. The time spent on observation corresponds to the delay time. Afterward, the video skips ahead by the delay duration, catching up to the live point. While Skip Playback instantly recovers delays, it disrupts temporal and viewpoint continuity.

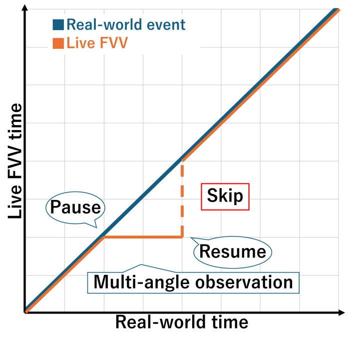

Figure 3: Changes in video-play time during skip playback. The horizontal axis represents real-world time, and the vertical axis represents the time of the live free-viewpoint video. The video-play skips an amount of time equivalent to the duration spent on observation.

Constant-Speed Playback

In Constant-Speed Playback, delays caused by observing key moments are recovered by high-speed playback at a fixed rate after resuming the video. Figures 4 (a) and (b) respectively depict the continuous changes in video-play time and the discrete changes in video-play speed during this process. This method maintains temporal continuity while recovering delays.

The specific processing steps are as follows. Upon resuming video-play, the time difference between the live point and the resumed point is obtained as the delay time Δt. The required duration T to catch up to the live point is calculated using Equation (1), where n is a constant.

$$T = n\Delta t \qquad (1)$$

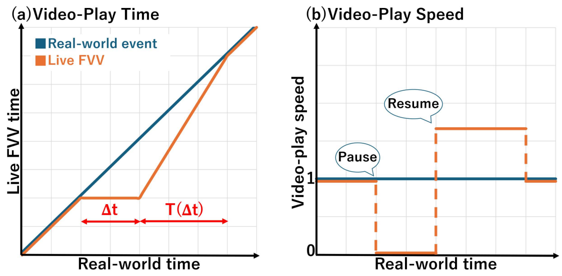

Figure 4: (a) Changes in video-play time during constant-speed playback. (b) Changes in video-play speed during constant-speed playback. The delay is recovered by high-speed playback at a fixed rate for a duration proportional to the delay time.

The video-play speed v needed to reach the live point within T is derived using Equation (2).

$$v = \frac{T + \Delta t}{T} = 1 + \frac{1}{n} \tag{2}$$

The video is played back at the calculated speed v for the duration T. Once catching-up with the real-world event, video-play returns to the normal speed of 1x. While Constant-Speed Playback preserves temporal and viewpoint continuity, the video-play speed transitions are discrete, switching between 0, v, and 1.

Accelerated-Decelerated Playback

In Accelerated-Decelerated Playback, video-play speed dynamically changes over time to recover delays. Figures 5 (a) and (b) respectively illustrate the continuous changes in video-play time and video-play speed during this process. Starting from 0, video-play speed gradually accelerates, reaches a maximum speed, and then decelerates back to normal speed upon the speed of the real-world events. This method maintains temporal, viewpoint, and video-play speed continuity.

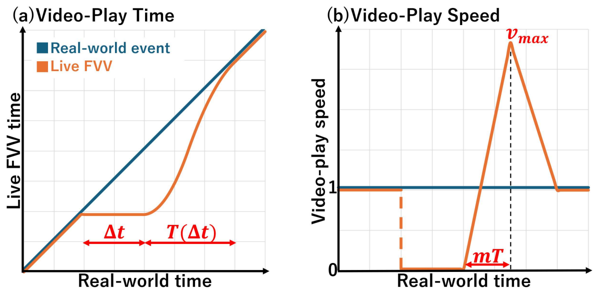

Figure 5: (a) Changes in video-play time during accelerated-decelerated playback. (b) Changes in video-play speed during accelerated-decelerated playback. The video-play speed after video resumption changes linearly before and after reaching the maximum speed.

The specific processing steps are as follows. As with Constant-Speed Playback, the delay time Δt is calculated, and the required duration T is determined using Equation (1). The time to reach the maximum speed from the start of catch-up playback is calculated as mT, where m is a constant greater than 0 and less than 1. Video-play speed changes linearly before and after reaching the maximum speed, as shown in Figure 5(b). The maximum speed $v_{\max}$ is determined based on the condition that the integral of video-play speed during catch-up playback matches the integral of real-world event progress. This relationship is expressed in Equation (3), and the derived maximum speed is given in Equation (4).

$$\Delta t + T = \frac{mTv_{max}}{2} + \frac{(1 + v_{max})(1 - m)T}{2} \quad (3)$$

$$v_{max} = 1 + m + \frac{n}{2} \quad (4)$$

EXPERIMENT

In order to evaluate the impression of three types of catch-up playback (Skip, Constant-Speed and Accelerated-Decelerated Playback) a questionnaire-based survey is conducted for comparing these methods. This section describes the experimental setup, procedure, evaluation methods, results, and discussions based on the findings.

Implementation

The video-play environment for free-viewpoint video (FVV) is developed using a VR environment, Unity. The implementation details are as follows. To ensure reproducibility of video playback during the evaluation, pre-recorded video captured from multiple viewpoints is prepared. The video-play system switches between viewpoints sequentially by simultaneously playing back all videos and toggling the displayed view. A pseudo-live video is used for catch-up playback, allowing continuous video-play without pausing. After pausing and switching viewpoints, delays relative to the pseudo-live video are recovered using the catch-up playback methods. For Constant-Speed Playback and Accelerated-Decelerated Playback, the duration T required to recover delays is set to 1.5 times the delay time Δt ($n = 1.5$). In Accelerated-Decelerated Playback, the time to reach maximum speed is set to half of T ($m = 0.5$).

Experimental Procedure

The experiment uses a soccer match video recorded from a bird's-eye view, as shown in Figure 6. The video is captured using 12 cameras arranged around the field. Three types of videos are prepared: those employing pause, viewpoint switching, and one of the three catch-up playback methods. These videos are shown to 15 participants. After viewing each video, participants evaluate their impressions using a 12-item, 7-point Semantic Differential (SD) method questionnaire. The evaluation items are shown in Table 1.

The evaluation uses a 7-point scale ranging from −3 to 3, with paired opposite adjectives as endpoints. The collected scores are analyzed using factor analysis to extract key evaluative factors. Additionally, the mean score for each evaluation item is calculated, and a t-test with a 95% confidence level is conducted to identify significant differences between the playback methods. This evaluation enables a comparison of the watching comfort provided by the three catch-up playback methods.

Figure 6: Soccer match video used in the experiment.

Table 1: Evaluation items used to evaluate video impressions.

Like	⇔	Dislike
Immersive	⇔	Non-immersive
Natural	⇔	Unnatural
Clean	⇔	Dirty
Pleasant	⇔	Unpleasant
Powerful	⇔	Weak
Calm	⇔	Rough
Relaxing	⇔	Stressful
Easy to watch	⇔	Hard to watch
Sense of speed	⇔	Lack of speed
Smooth	⇔	Jittery
Clear	⇔	Blurry

RESULTS

The factors are determined based on eigenvalues of at least 1 estimated through factor analysis, resulting in two factors. Additionally, factor loadings are calculated to indicate the extent to which the two factors are reflected in each of the 12 questionnaire items. The results of the factor loadings estimation are presented in Table 2. The principal factor method is used for estimating the factor loadings, and varimax rotation is applied for factor rotation. Factor analysis identifies two primary factors: the first factor, including items like "natural," "easy to watch," and "pleasant," is defined as Comfort, while the second factor, including items like "powerful," "sense of speed," and "immersive," is defined as Dynamism. The evaluation of the videos is thus found to be based on these two axes: Comfort and Dynamism.

Table 2: Results of factor loadings estimation. Items with an absolute factor loading of 0.5 or higher for the first factor are highlighted in red, while those with an absolute factor loading of 0.5 or higher for the second factor are highlighted in blue.

Evaluation Items	First Factor	Second Factor
Like	0.74	0.33
Immersive	0.33	0.75
Natural	0.84	0.07
Clean	0.71	0.02
Pleasant	0.77	0.04
Powerful	0.09	0.83
Calm	0.61	−0.70
Relaxing	0.77	0.04
Easy to watch	0.76	0.49
Sense of speed	−0.07	0.88
Smooth	0.35	0.57
Clear	−0.02	−0.18

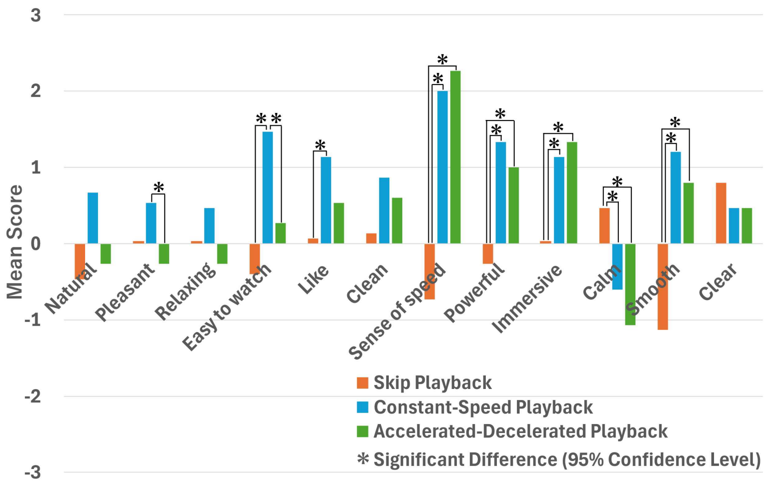

Figure 7: Comparison of the mean scores for each evaluation item across three catch-up playback methods. Items showing significant differences between methods at the 95% confidence level are marked with an asterisk "*."

Figure 7 presents the average scores for each evaluation item across three catch-up playback methods as a bar graph. Among the 12 evaluation items, the discontinuous playback method (Skip Playback) scores lower than the continuous playback methods (Constant-Speed Playback and Accelerated-Decelerated Playback) on 8 items, indicating that continuous playback is generally preferred over discontinuous playback. Additionally, Figure 7 highlights items with significant differences between playback methods, marked with "*," based on the t-test. Constant-Speed Playback scores significantly higher than Skip Playback for "easy to watch" and "like." For "easy to watch," Accelerated-Decelerated Playback scores significantly lower than Constant-Speed Playback. Regarding Comfort, Constant-Speed Playback receives the highest rating, while Skip Playback scores the lowest.

For Dynamism, both Constant-Speed Playback and Accelerated-Decelerated Playback are rated higher than Skip Playback.

DISCUSSION

The results indicate that continuous playback methods (Constant-Speed Playback and Accelerated-Decelerated Playback) receive higher ratings than the discontinuous method (Skip Playback). This finding suggests that viewers tend to prefer video content with temporal continuity. Skip Playback, which skips parts of the video, appears to disrupt the visual continuity, potentially negatively impacting viewers' experiences. Among the three methods, Constant-Speed Playback receives the highest ratings for Comfort, suggesting that it is the most effective method for recovering delays in live FVV without compromising watching comfort. The lower rating of Accelerated-Decelerated Playback for "easy to watch" suggests that continuous changes in video-play speed during catch-up playback may reduce watching comfort. Dynamic changes in video-play speed may cause visual discomfort, increasing the cognitive load for viewers.

These findings indicate that maintaining temporal continuity is essential for watching comfort, while video-play speed continuity is not necessarily as critical.

CONCLUSION

This study proposed a catch-up playback method to address delays in live free-viewpoint video caused by multi-directional observation of key moments. Specifically, three types of playback methods—Skip Playback, Constant-Speed Playback, and Accelerated-Decelerated Playback—were examined, and their effects on viewers were compared and evaluated. The experimental results demonstrated that Constant-Speed Playback is the most effective method for recovering delays while maintaining watching comfort. Furthermore, the findings revealed that temporal continuity in video-play plays a crucial role in ensuring watching comfort.

REFERENCES

Amar, I., El-Sana, A. & El-Sana, J., 2023. Live Free-View Video for Soccer Games. *Web3D '23: Proceedings of the 28th International ACM Conference on 3D Web Technology,* Issue 15, pp. 1–6.

Hertzog, C., Sakurai, S., Hirota, K. & Nojima, T., 2020. Toward Augmented Reality Displays for Sports Spectators: A Preliminary Study. *Proceedings of The 13th Conference of the International Sports Engineering Association,* 49(1), p. 129.

Kıyak, Y. S., Budakoglu, I.. I., Masters, K. & Coskun, O., 2023. *Medical Teacher,* 45(8), pp. 913–917.

Kitaguchi, D., Kumano, K., Takatsuki, R., Xie, C., Hashimoto, S., Akashi, Y., Kitahara, I., & Oda, T., 2024. Free-Viewpoint Video in Open Surgery: Development of Surgical Arena 360. *Journal of Surgical Education,* 81(3), pp. 326–329.

Koyama, T., Kitahara, I. & Ohta, Y., 2003. Live mixed-reality 3D video in soccer stadium. *Proceedings of The Second IEEE and ACM International Symposium on Mixed and Augmented Reality,* pp. 178–187.

Mo, C.-Y., Wang, C., Dai, J. & Jin, P., 2022. Video playback speed influence on learning effect from the perspective of personalized adaptive learning: A study based on cognitive load theory. *Frontiers in Psychology,* 13(839982).

Nagai, T., Shishido, H., Kameda, Y. & Kitahara, I., 2018. An On-site Visual Feedback Method Using Bullet-Time Video. *Proceedings of ACM MM2018 First International Workshop on Multimedia Content Analysis in Sports,* pp. 39–44.

Nonaka, K., Watanabe, R., Chen, J., Sabirin, H. & Naito, S., 2018. Fast plane-based free-viewpoint synthesis for real-time live streaming. *Proceedings of 2018 IEEE Visual Communications and Image Processing,* pp. 1–4.

Perez, P., Corregidor, D., Garrido, E., Benito, I., Gonzalez-Sosa, E., Cabrera, J., Berjon, D., Diaz, C., Moran, F., Garcia, N., Igual, J & Ruiz, J., 2022. Live Free-Viewpoint Video in Immersive Media Production Over 5G Networks. *IEEE Transactions on Broadcasting,* 68(2), pp. 439–450.

PwC, 2024. *Streaming the game: How the rise of digital platforms is changing sports consumption.* [Online] Available at: https://www.pwc.com/us/en/industries/tmt/library/sports-streaming-platforms.html [Accessed 8 1 2025].

Sankoh, H. Naito, S., Harada, M., Sakata, T & Minoh. M., 2015. Free-viewpoint Video Synthesis for Sport Scenes Captured with a Single Moving Camera. *ITE Transactions on Media Technology and Applications,* 3(1), pp. 48–57.

Smolic, A. Mueller, K., Merkle, P., Rein, M. Kautzner, T., Eisert, P & Wiegand, T., 2004. Free viewpoint video extraction, representation, coding, and rendering. *Proceedings of the ICIP 2004, IEEE International Conference on Image Processing.*