2022 IEEE on Conference Virtual Reality and 3D User Interfaces (VR) | 978-1-6654-9617-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/VR51125.2022.00017

Bullet Comments for 360° Video

Yi-Jun Li[1] Jinchuan Shi[1] Fang-Lue Zhang[2] Miao Wang[1,3*]

1 State Key Laboratory of Virtual Reality Technology and Systems, SCSE, Beihang University, China
2 Victoria University of Wellington, New Zealand 3 Peng Cheng Laboratory, China

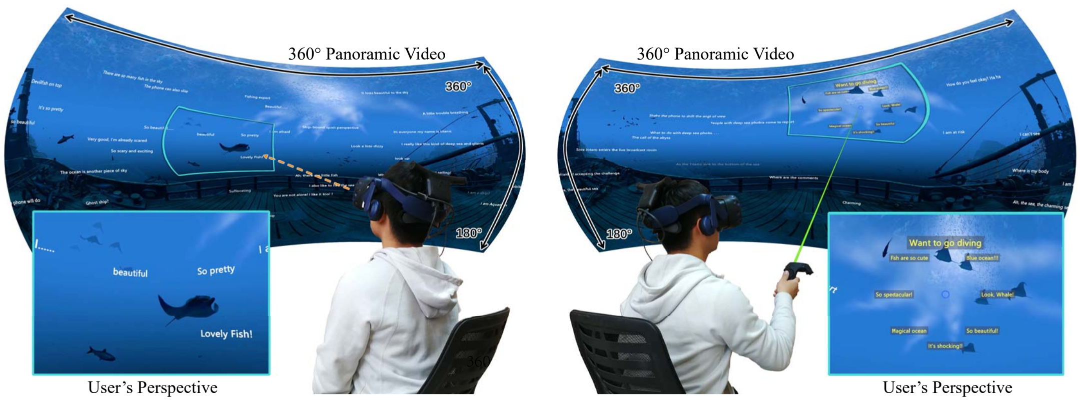


Figure 1: Illustration of bullet comments for 360° video: the user is sitting in a swivel chair and experiencing a 360° video via HMD and controller. **Left:** the user can see the bullet comments within the field-of-view, displayed on a virtual spherical surface formed by the 360° video. **Right:** the user inserts a comment to the 360° video via controller using our *Locate-and-Select* method.

Abstract

Time-anchored on-screen comments, as known as *bullet comments*, are a popular feature for online video streaming. Bullet comments reflect audiences' feelings and opinions at specific video timings, which have been shown to be beneficial to video content understanding and social connection level. In this paper, we for the first time investigate the problem of bullet comment display and insertion for 360° video via head-mounted display and controller. We design four bullet comment display methods and evaluate their effects on 360° video experiences. We further propose two controller-based methods for bullet comment insertion. Combining the display and insertion methods, the user can experience 360° videos with bullet comments, and interactively post new ones by selecting among existing comments. User study results revealed how the factors of display and insertion methods affect 360° video experience. With the experiment findings, we also discuss useful design insights for 360° video bullet comments.

1 Introduction

With the rapid development of the internet and social media, online video streaming experiences have been immensely enriched by facilitating computer-mediated interaction between audiences of the same video. Compared to watching videos passively, audiences have been increasingly demonstrating their desire for interaction with peers in a social environment when watching online videos [39]. *Bullet comments*, also known as Danmaku in Japanese and Dan Mu in Mandarin, are a technology that enables audiences to post live time-anchored comments that float over the playing video to share their emotions and thoughts with other audiences. They have been an appealing feature in social media platforms such as YouTube (with DMOOJI plugin), Niconico, and Bilibili, which can improve audiences' comprehension of video content and strengthen the social connections between individuals [21].

Due to the availability and advancement of commodity omnidirectional cameras and 360° video processing and streaming technology, viewing online panoramic video via head-mounted displays (HMDs) has been attracting more public interests than a couple years ago. Compared to the traditional narrow field-of-view (FoV) videos, 360° videos provide the audience a unique experience that is more immersive and diversified [2,16]. A 360° video is recorded in all directions, usually stored in equirectangular format, and can be projected onto a sphere to offer a complete 360° view. Users are allowed to sense the action and explore visual content from all viewing directions. Although some in-headset collaborative viewing and interaction methods [31,42,43] have been proposed to enhance the sociability of 360° video, they cannot generalize to support online streaming applications due to the fragility of their visualization and communication methods when dealing with large real-time data. Another problem of these methods is that they require all users to attend simultaneously for online communication, while time-anchored comments have an advantage of asynchronous communications, so that the users can watch videos and get the feeling of social connections at their convenient times.

To add more interactivity and sociability to the online 360° video viewing experiences, we propose to introduce the bullet comment technique to 360° video streaming, and investigate how bullet commenting should look like for 360° videos. Specifically, we look into the in-headset display and insertion methods of bullet comments for 360° videos. Intuitively, a live comment should be displayed right after posted by floating over the viewport that is spatially consistent with the commenting position. Moreover, considering that the audiences of 360° videos can freely navigate their viewports and perform controller-based interactions, we believe that the display of the comments needs to be time-space-anchored, and the interaction method enabling bullet comment insertion has to be accurate and convenient when performing by VR headset and controller.

We thus raise two main research questions (RQs) associated to bullet comment display and interaction techniques for 360° videos:

*Corresponding Author: miaow@buaa.edu.cn

978-1-6654-9617-9/22/$31.00 ©2022 IEEE
DOI 10.1109/VR51125.2022.00017

RQ1: What is the most popular approach to display bullet comments on 360° videos for audiences?

RQ2: Which bullet comment insertion method will audiences find more natural and easier to use?

To answer these questions, we designed bullet comment display and insertion methods for immersive 360° video. We conducted two user studies to discuss the answers in depth and offer possible insights. To the best of our knowledge, this work makes the first attempt to explore bullet comments for immersive 360° video experienced with VR headset and controller.

2 Related Work

2.1 Panoramic Image and Video Processing

The prevalence of low-cost 360° cameras and display devices has been making watching 360° videos in the immersive virtual environment more popular among the general public. Other than traditional narrow FoV videos, panoramic videos can provide a more immersive experience by allowing audiences to continuously explore the video content from any perspective [29]. These videos can be further processed for better visual effects. For example, they might be tilted when the camera is not upright, which should be corrected by upright adjustment methods for better visual comfort [14, 15]. Jung et al. [14] presented an automatic method for upright adjustment which minimizes the cost function formulated by horizontal and vertical lines in the scene. A deep learning-based method was also proposed to estimate the camera orientation and return its upright version [15]. The quality of panoramic videos sometimes suffers from shakiness when the camera is moving and needs to be improved by stabilization [18, 19, 44, 49]. Kopf's method [18] estimated key frame rotations by 3D analysis and maximized the visual smoothness by 2D optimization, which can remove small jitters and parallax from the video. Panoramic videos can also be used in mixed reality interaction. Rhee et al. [33] proposed the MR360 system that can synthesize the light source from a low dynamic range 360° video to illuminate the virtual objects in real time. Tarko et al. [45] proposed a new approach to reconstruct camera motion, which allows users to insert virtual objects in a dynamic panoramic video. Wang et al. [47] proposed a ray casting based bidirectional shadow rendering method to generate shadows of virtual objects in real time. Panoramic videos can also provide audiences with immersive live streaming experiences, but they are extremely bandwidth intensive to achieve acceptable visual quality. Since most of the pixels in a panoramic video would be out of the users' views, viewport-adaptive streaming methods [10, 11, 22, 50] were proposed to lower the bandwidth cost for online streaming. For example, Hosseini et al. [11] proposed a divide and conquer approach by using viewport adaption techniques, which can significantly reduce the bandwidth to deliver high-quality videos. The attempts of adding motion parallax to panoramic videos can be beneficial to enhance immersion [1, 13, 26, 37]. Other approaches of saliency detection for panoramic videos can help audiences focus on the most interesting content and improve the efficiency of exploration [5, 27, 30, 56].

2.2 360° Video Exploration

Although 360° video offers audiences the freedom to select the target view based on their interests [57], sometimes it can be tiring and annoying when the scene is complex. Automatic view navigation can relieve audiences from constantly looking for interesting regions, which are mainly focused on how to generate a normal FoV camera path to follow the most valuable region [12, 16, 40, 48]. Kang and Cho [16] proposed a navigation system which utilizes saliency and optical flow to iteratively calculate the best NFoV camera movement, making the camera smoothly follow the most interesting object. Collaborative viewing experiences have also been investigated by virtually teleporting the remote user to the place of the local user [20, 34, 53]. Rhee et al. [34] proposed an asymmetric platform by placing a 360° camera for the local user to capture the surroundings and send them to the remote VR user's HMD, enabling the remote user to interact in the same real space. For collaborative panoramic video editing, Nguyen et al. [31] proposed the CollaVR system to support multiple users reviewing and communicating in their VR headsets. Chang and Cohen [3] proposed a novel interface to enable users to magnify certain regions of high-resolution panoramas to explore more details, and alleviate the motion sickness to improve the user experience at the same time. Some approaches use visual guidance to draw users' attention to certain areas of panoramic images or videos [36, 46]. Lin et al. [25] proposed the Outside-In interface which indicates the off-screen region-of-interests (ROIs) as floating thumbnail windows, making the spatial relationship between the current view window and the ROIs more understandable to realize a more efficient navigation guidance. Wallgrün et al. [46] compared three visual guiding mechanisms (arrow, butterfly guide, and radar) and concluded that the arrow mechanism was the most favored approach. Yamaguchi et al. [51] presented a novel technique utilizing a hand-held thumbnail to offer the full 360° reference to the VR user, which can help relieve the fear of missing out.

2.3 Bullet Comments

Bullet comments are a kind of time-anchored live comments which are popular in online video streaming applications, especially in eastern Asia [7, 8, 52]. When the function is enabled, the posted comments will float across the playing video in real-time and are visible to all the audiences [7]. If there are a lot of comments for a certain time of the video, comments will cover the entire screen, making bullet comments into 'bullet screens' [8]. Compared to traditional online videos watching experiences, bullet comments can make videos more understandable [6] and potentially enhance social interactions [21], which can bring great opportunities and benefits to the promotion of their services [7]. Besides, bullet comments can also reflect the sentiment and attitude of audience towards the video [6], which can be utilized for emotion analysis [6, 23, 35]. Cui et al. [6] proposed a novel scheme based on emoticons and tone texts, improving the analysis of bullet comment sentiment. For online learning, Lee et al. [21] proposed a time-anchored commenting interface which allows students to exchange comments on the same video clip, where comments will appear at corresponding timings during playback. It revealed that time-anchored commenting can enhance the learning engagement and the dynamically display method can enhance social interactivity. Liao et al. [24] constructed virtual classmates by synthesizing previous learners' time-anchored comments, which can help learners have the sense of being accompanied and achieve better learning outcomes. Some work makes use of bullet comments to summarize videos or films to provide fast primary impressions to audiences [4, 41, 54]. Xu et al. [4] combined video frames with time-synchronized comments to encode various user preferences in a unified multi-modal space for key frame recommendation to help video preview or retrieval. Sun et al. [41] proposed a novel approach for highlight detection and summarization in movies using bullet comments and relevant visual information. In video sites that support 360° video playing, such as Bilibili, the spherical video frames are projected to a rectilinear display window. All the bullet comments are just displayed on the video player's window as what is done for 2D videos, regardless of what the user is viewing. No previous work has investigated the appropriate approaches to display and insert bullet comments on 360° videos.

3 Bullet Comment Display and Insertion

Following the research questions in Section 1, our study focuses on two major aspects: how to show bullet comments to audiences and how the audiences insert their live comments conveniently. Just like normal 2D videos, bullet comments for spherically projected 360° videos should be easy to read and input. Here, we discuss the design

(a) Planar-Horizontal (b) Spherical-Static (c) Spherical-Horizontal (d) Spherical-Vertical

Figure 2: Illustration of bullet comment display methods. The arrows and ghost effects are added for illustration to indicate the moving directions of comments on the screen.

criteria and present our proposals for bullet comment display and insertion.

3.1 Design Criteria

Exploring 360° videos is a significantly different experience compared with normal 2D videos: 360° video has omnidirectional FoVs, where a user can actively view the content of a selected partial FoV in an HMD by rotating the head and perform interactions using controllers. When extending bullet comment display and insertion methods from 2D videos to 360° video, the following design criteria (DC) need to be considered:

DC1: ***Time-space-anchored.*** In addition to aligning with the commenting time, a live comment should be visualized spatially consistent with the corresponding visual content. Otherwise, users can be distracted and confused by the comments irrelevant to what they are watching from their current viewport.

DC2: ***Simple, accurate and fast.*** To maintain a smooth viewing experience, the operations for comment posting should be as easy as possible so that the user can perform fast and accurate interaction with no need to pause the video.

3.2 Bullet Comment Display

DC1 implies that the posted comments should be shown around what the audience is watching when posting the comment, and keep them for a short period right after the commenting time. In this way, the user will be only exposed to the comments describing the content related to the current FoV. Based on this criterion, we present four time-space-anchored methods and will evaluate each of them later:

***Planar-Horizontal (PH)*:** Bullet comments are displayed on a virtual planar canvas attached to the whole FoV of the user. Once the user's FoV covers the position of a time-space-anchored comment, the comment enters the user's FoV from the right boundary and slides evenly and horizontally to the left boundary within a 20-second period, catering to common reading habits. It simulates the movement of bullet comments used by current video streaming websites.

***Spherical-Static (SS)*:** Each time-space-anchored comment appears at its insertion position on the spherical surface formed by 360° video, starting from the insertion timestamp. Similar to *PH*, the comment disappears after a 20-second period. Fade-in and fade-out effects are used on the texts for smooth transitions.

***Spherical-Horizontal (SH)*:** Each comment appears at its insertion time and position on the sphere. Moreover, the comment horizontally and counter-clockwise flies along the latitude from the insertion position for 20 seconds, with a 6°/s angular velocity.

***Spherical-Vertical (SV)*:** Each comment appears at its insertion time and position on the sphere, and flies upward along the longitude from the insertion position with a 6°/s angular velocity for 20 seconds or till reaching the pole region.

Figure 2 illustrates the display methods. In all the implementations of the above display methods, following the appearances of conventional bullet comments, we empirically set the font color as white and the typeface as Arial, respectively. The font size was adjusted to be visually moderate (80 pt height) and consistent across methods. We determined the comment density via a 5-subject pilot study, where moderate density (120 comments per minute) was more preferred than sparse (30 comments per minute) and dense (600 comments per minute).

3.3 Bullet Comment Insertion

Posting a comment on a 360° video typically needs two steps, comment creation and insertion. Comment creation involves character input in virtual environment, which is an important and general research topic. There are various character/text input methods that can be employed, such as virtual keyboard or speech-to-text. Since how to input characters in VR is out of the scope of our research, we simplify the comment input step by providing candidate comment texts to users and just let them to insert comments using controller. More specifically, for each 360° video, we provide a candidate list with 8 pre-set comments related to the scene (see supplementary materials). We investigate how to select and insert a comment from the given comment list by using a single hand-held controller and propose two methods that reflect **DC2** (see Figure 3):

***Select-and-Drag (SD)*:** the user activates the interface where the candidate comments are circularly distributed by clicking the trackpad, then selects a comment by touching the corresponding trackpad region. After selection, the user drags the comment to the target position on the video by holding the trackpad, and inserts the comment by releasing the trackpad.

***Locate-and-Select (LS)*:** the user first points at the target insertion position on the viewport using raycasting, then clicks the trackpad to activate the comment insertion interface. After that, the user selects a comment by clicking the corresponding trackpad region. Then the selected comment is immediately inserted at the target position.

To distinguish the candidate comments and existing bullet comments in the video, the candidate comments are illustrated in yellow color over a gray rectangle. When selecting comments, the currently selected one is highlighted in a bigger font size. Operations such as pausing the video by pressing the controller's trigger, or revoking the last comment by pressing the controller's menu button are supported to provide better flexibility to users.

4 Study 1: Bullet Comment Display

In this study, we collect 360° videos with bullet comments from the internet, and investigate users' behavior and subjective perception over different bullet comment display methods.

4.1 Participants, Apparatus, and Materials

We recruited 20 participants (9 female, 11 male), aging from 19 to 27 (Mean = 23.20, SD = 1.85). 2 of them had no VR experience before, and 12 had experienced VR applications for less than five times. All the participants had normal or corrected normal vision and had watched 2D normal videos with bullet comments before, but half of them had no 360° video experiences.

The experiment was conducted in a lab with a 4m × 4m physical tracking space. An HTC Vive Pro Eye HMD was used to provide a 2880 × 1600 combined resolution (110° FoV) at a refresh rate of 90 Hz, as well as the audio. The 360° video playing system was implemented in Unity and ran on a PC with an Intel Core i7

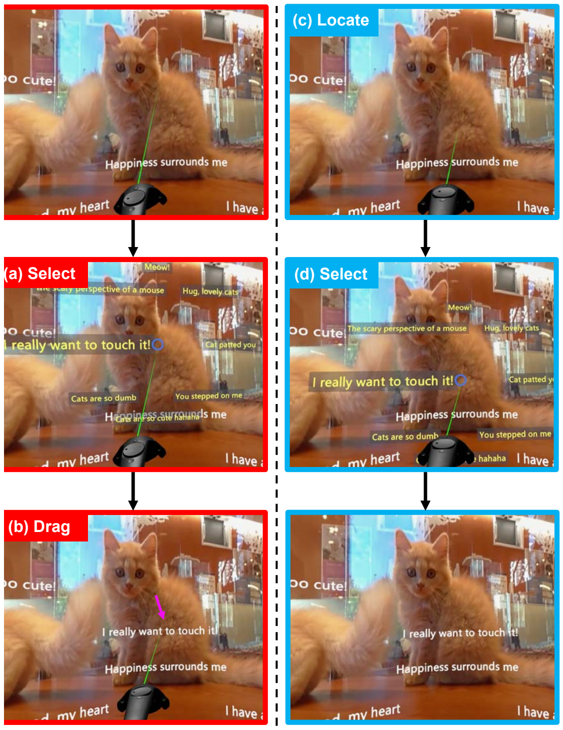


Figure 3: Illustration of comment insertion methods. **Left**: *Select-and-Drag*. (a) and (b) are the screen captures when a user is selecting and dragging a comment. **Right**: *Locate-and-Select*. (c) and (d) are the screen captures when a user is locating a target position and selecting a comment.

processor, 32GB RAM and a GeForce RTX 2080Ti GPU. To offer a comfortable viewing experience, participants sat in a swivel chair, free to rotate the head and body to fully explore 360° videos.

We collected five 360° videos (see Figure 4) with corresponding time-anchored comments in the local language from one of the most popular online video sites Bilibili, and cropped the videos to 1.5 to 2 minutes. The virtual camera for each video remains stationary, which ensures the content stable and comfortable for watching. Similar to previous work [28, 38], we manually labeled the region of interest (ROI) using rectangles to represent the most interesting events (e.g., cats jumping or whale swimming). Moreover, since only timestamp information was available for each downloaded comment, we further manually adjust its spatial position to ensure the alignment between the visual content and the comment. We clarify that original local-language bullet comments were used in our studies. The English bullet comments illustrated in this manuscript and shown in the supplementary video were post-processed by Google Translate service to show the comments in English for a better illustration. For more details about the dataset, please refer to our supplementary materials.

4.2 Hypotheses and Conditions

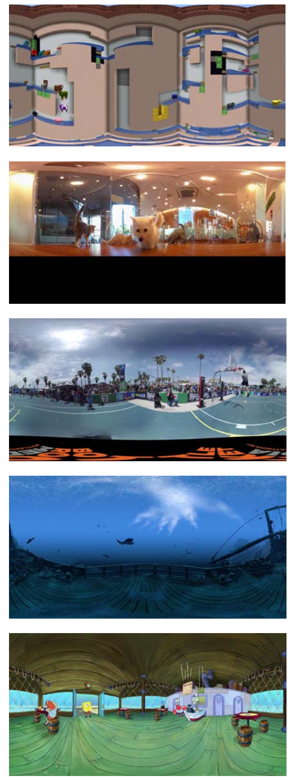
As aforementioned, bullet comment display methods *PH*, *SS*, *SH*, and *SV* are evaluated. We make the following main hypotheses:

H1: Spherical display methods *SS*, *SH*, and *SV* are preferred to the planar display method *PH*.

H2: *SH* is the most favorite method among spherical display methods.

4.3 Design and Procedure

We design a within-subject user study to test the above hypothesis. We let subjects watch 360° videos equipped with bullet comments,



Sheep Sports Game: A racing game where a flock of sheep are racing in a maze with many traps blocking them.

Cats Surrounding: A video clip of a pet store where four lovely kittens playing with each other.

Dunk Show: A show of fancy dunks by professional basketball players on the Venice Beach basketball court.

The Blu: A video clip captured in deep water. Audiences can feel like standing on the deck of a sunken ship and closely contacting the largest species on Earth.

Inside the Krusty Krab: A fragment of the life in Bikini Bottom with the whole Krusty Crew inside the Krusty Krab.

Figure 4: 360° videos used in the user studies.

record their behaviors and collect subjective questionnaires.

When each participant arrived at the lab, we first gave an information sheet about the goal and process of the experiment. The participant read it and signed a consent form. Then the participant was asked to fill in a short demographic survey. We taught the participant how to wear the HMD and calibrate the HMD eye tracking module. To help the participant better understand the functions of different bullet comment methods, each display method was experienced on the same *Sheep Sports Game* video in a training session. After completing all the four training trials for the different methods, the participant took four formal trials, where each trial used one of the bullet comment display methods on a selected 360° video. The order of bullet comment methods across participants was counterbalanced using a Latin Square approach and the order of videos (excluding the training video) was also randomized. A 2-minutes short break was allowed between trials. The participant was asked to fill in questionnaires before and after each formal trial. Finally, a short interview was conducted to gather free-form subjective feedback. Each participant spent 22 minutes on average and was finally awarded a gift.

Regarding measures, we recorded the user behavior from the eye tracking module and collected subjective feedback through questionnaires and face-to-face interviews. Specifically, we recorded participants' visual fixation (VF) inside ROIs and on comments, measured as the percentage of frames that gazing at ROIs and bullet comments. Moreover, we recorded gaze movement as the average angular changes of gazes during the experiment. A pre-SSQ questionnaire [17] was collected before each formal trial and the post-SSQ, engagement [32] and social interaction questionnaires [55, 58] were collected after the trial. Finally, an additional 7-Likert scale questionnaire was collected to rate the user preference for each bullet comment display method, and open comments were collected from the interview.

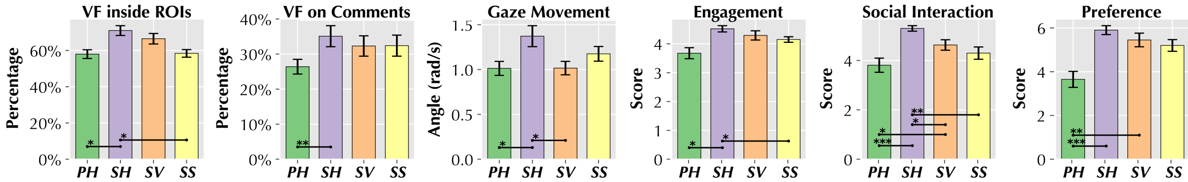


Figure 5: Means and standard errors of measures in Study 1. Significant differences are indicated with $*p<0.05$, $**p<0.01$, $***p<0.001$.

4.4 Results

In total, we collected data from 80 valid trials (20 participants × 4 trials). Kolmogorov-Smirnov tests revealed normally distributed data for most measures except for *simulator sickness* and *user preference*. Therefore, repeated measures ANOVA (RM-ANOVA) tests at the 5% significance level and Bonferroni-adjusted post-hoc tests were adopted. Greenhouse-Geisser corrections were performed if the sphericity assumption was violated. Figure 5 shows the means and standard errors of measures of display methods. For the non-normally distributed *simulator sickness* and *user preference* data, we used non-parametric Friedman tests to analyze the main effect. When the effect was significant, we conducted Wilcoxon signed-rank tests with Bonferrorni correction for post-hoc analysis.

Visual fixation inside ROIs. We found that *display method* had a significant effect ($F(3,57)=6.356, p=.001, \eta^2=.251$) on this measure. *SH* had the largest value and was significantly larger than *PH* ($+.130, p=.039$) and *SS* ($+.126, p=.017$).

Visual fixation on comments. A significant effect of *display method* on this measure was found ($F(3,57)=3.656, p=.018, \eta^2=.161$). Although *SH* held the largest value, it only had a large difference with *PH* ($+.087, p=.005$).

Gaze movement. *Display method* had a significant effect ($F(3,57)=6.834, p=.001, \eta^2=.265$) on gaze movement. *PH* and *SS* got very low gaze movement values, and were significantly lower than *SH* ($-.359, p=.013$ for *PH*, and $-.356, p=.028$ for *SS* respectively).

Engagement. *Display method* was observed a significant effect on user engagement score ($F(3,57)=5.868, p<.001, \eta^2=.236$). *SH* and *SV* had the top two scores, and the results revealed that *SH* was significantly larger than *PH* ($+.846, p=.014$) and *SS* ($+.371, p=.030$).

Social interaction. A significant effect of *display method* on social interaction was indicated by the result ($F(3,57)=12.928, p<.001, \eta^2=.405$). *SH* got the largest score and it has significant difference with *PH* ($+1.490, p<.001$), *SV* ($+.670, p=.048$) and *SS* ($+1.000, p=.004$). In addition, *SV* was significantly larger than *PH* ($+.820, p=.043$).

User preference. *Display method* was observed a significant effect on the user preference ($\chi^2(3)=16.541, p<.001$). *SH* (Mean = 5.90) and *SV* (Mean = 5.45) achieved the best and second best ratings, and were significantly larger than *PH* (Mean = 3.65) ($p<.001$ for *SH* and $p=.007$ for *SV*). *SS* (Mean = 5.20) was larger than *PH*, but the difference was not significant ($p=.120$).

Simulator sickness. A Kolmogorov-Smirnov test indicated that the SSQ data was not normally distributed. Thus, we used a Wilcoxon signed-rank test at the 5% significance to analyze the differences between the scores of pre-SSQ and post-SSQ. The results showed that there was no significant difference between them, although pre-SSQ scores (Mean = 9.26) were always relatively lower than the post-SSQ scores (Mean = 12.76).

4.5 Interview Findings

We gathered open comments for further analysis of the effects of bullet comment display methods on 360° video experience. The findings can be summarized in three major aspects:

Benefits and shortcomings of bullet comment display. Except for one participant that thought bullet comments were useless, all the others regarded bullet comments as a useful technique to help them understand the 360° video content better: *"Some floating comments guided me to notice the important events outside my field of view, such as the swimming whale. (P1)" "Bullet comments offered some supplementary information and knowledge to help me understand the video. (P14)"* Moreover, bullet comments can also enhance social connections: *"Other guys' comments were funny. (P17)" "I had great empathy with the other audiences when watching 360° video, feeling like I was not watching this video alone. (P6, P11)"* Despite the benefits, participants also complained about the shortcomings: *"Some comments were meaningless or rude, which made me not feeling good. (P7)" "Sometimes I would pay more attention to the comments rather than the video itself. (P20)"* Some participants made suggestions to improve the bullet comment display methods: *"Comments can be translucent, which will alleviate the occlusion. (P20)" "Other colors can be used in bullet comments to make them more interesting and salient especially when they overlap with a white background. (P5)" "The foreground objects should not be occluded by the comments. (P17)"*

Differences between planar and spherical display methods. Overall, spherical display methods, *SS*, *SH*, *SV* were all preferred over the planar method *PH* due to the visual compatibility with 360° videos and a higher degree of immersion. *"SH was consistent with the normal reading habit of text from left to right, and provided a better immersive feeling than PH. (P18)"* The major problem of *PH* was about the discomfort from the broken immersion: *"Comments were not well embedded within the 360° video, which made me not able to focus on both the comments and video content at the same time. (P5)" "The floating text seemed too close to my eyes, which put great pressure on me. (P12)" "The comments followed my viewport all the time, which made me uncomfortable and dizzy when I move or rotate my head. (P9)"*

Comparison among spherical methods. *SH* was acknowledged as the most favourite method. One possible reason could be that its similar movement of 360° bullet comments with the traditional bullet comment display method in 2D made participants feel more familiar. Moreover, both *SH* and *SV* were better than *SS* because flying texts facilitate the reading of those comments: *"I like SH because it looks natural and I can read the comments with no effort. (P18)" "Comments in SS were closely related to the content around them, but sometimes the comments can occlude an interesting region for a long time. (P13, P19)" "When experiencing SS, I have to change my viewing direction more frequently to look for new comments. (P12)"* Despite the overall positive feedback, some shortcomings for SH and SV were reported: *"I have to move my head to follow the comments' movements when I want to read the text. (P14)"* Only two participants liked *SS* better because the static display method did not deliver them any dizziness.

4.6 Discussion

Participants preferred spherical display methods to the planar one, given the evidence of significantly larger *user preference* values of *SH* and *SV* than *PH*. The two subjective measures, *engagement* and *social interaction*, further confirmed this conclusion. According to the interviews, a possible reason is that spherically displayed bullet comments conform to the visual features of 360° videos, providing users a higher level of immersion than the planar method. On the contrary, some participants argued that migrating a 2D planar display method to virtual environments can make them dizzy, stressed and distracted, but our SSQ results showed this is an individual phenomenon (there was no significant difference between the pre-SSQ scores and the post-SSQ scores for each display method). These disadvantages further degrade the overall viewing experience (lower *engagement* and *social interaction* scores), make users difficult to find important targets (lower *VF inside ROIs*) and sort of unwilling to watch bullet comments (lower *VF on comments*) as well as reduce the exploration desire (lower *gaze movement*). Therefore, we conclude that hypothesis **H1** can be accepted.

Regarding the three spherical display methods *SS*, *SH* and *SV*, *SH* was the favorite one with the largest *engagement* and *social interaction* scores, where most of the differences with alternatives were significant, except for the *engagement* of *SH* over *SV*. The records of user behavior also showed that when experiencing *SH*, participants found it easier to locate at interesting targets (highest *VF inside ROIs*, significantly higher than *SS*). Participants were more willing to explore the content with *SH* (largest *gaze movement*, significantly larger than *SV*) and read bullet comments (highest *VF on comments*). One possible explanation could be that the movement of *SH* is similar to the traditional bullet comment display method and all the participants had seen bullet comments before, making *SH* easier to be accepted by participants. Both *SH* and *SV* won higher *user preference* and *social interaction* than *SS* because participants can receive more information from sliding bullet comments without the need of changing gaze direction. According to the participants' descriptions, *SS* led to larger *gaze movements* because participants preferred to search for bullet comments by changing visual fixations (although we did not ask them to do so), which was consistent with the lower *VF inside ROIs* value of *SS*. Therefore, the above findings validate hypothesis **H2**.

According to the interview, almost all the participants confirmed that bullet comments were helpful for understanding 360° video content, strengthening social connections and improving enjoyment.

Design Insights. We distill design insights for comment display on 360° videos from the experiment results as follows:

DI1: Displaying comments on a spherical surface can increase the immersion. Based on the previous findings, embedding bullet comments on the spherical surface of the 360° video can provide users with a more immersive experience and strengthen the engagement and social interaction. This approach can also reduce simulator sickness to improve user comfort. The side effects of planar display including dizziness, pressure and distraction, should be avoided when designing new comment display methods.

DI2: Sliding texts provide a better user experience. In our experiment, sliding comments horizontally or vertically (*SH*/*SV*) won more user preference than the static ones (*SS*/*PH*), although there was no statistically significant difference among them. Interviews also indicated that sliding comments were more vivid, providing more information, and reducing the feeling of content obscuring, which could further enhance engagement and social interaction.

Moreover, our pilot study indicated that the designer should also consider the trade-off between comment density and video content cleanness, since sparse comments may be less enjoyable while the dense comments can confuse users.

Note that we do not set "no comments" as baseline based on the assumption that bullet comments would provide better user experiences for users as shown in the previous works [6, 7, 21]. In addition, as the comments are collected from 2D video website, many of them may not have strong spatial linkage with certain content, which might count against *SS*. Due to the bias of bullet comments' spatial meaning, the comparison between *SH*/*SV* and *SS* might be unfair for some videos.

5 Study 2: Combined Display and Insertion

The results from the first study revealed the benefits and shortcomings of bullet comment display methods. To answer **RQ2**, we further investigate different comment insertion methods using VR headsets based on our findings in bullet comment display.

As aforementioned in Section 3, to reduce the complexity and uncertainty during comment insertion, we focus on how to select a live comment from a given candidate comment list and insert it at a desired position using a single hand-held controller. We combine the two proposed methods *SD* and *LS* with the winners of comment display methods so that users can have a full experience by watching existing bullet comments and inserting new comments. Specifically, we pick up the best and the second best display methods *SH* and *SV*, forming a 2 (display) $\times$ 2 (insertion) condition combination. We derived two hypotheses in this section:

H3: *LS* is preferred by users. We hypothesize this because the operation of *LS* is expected to be simpler and faster than *SD*.

H4: The performance of insertion methods would not be significantly affected by display methods.

5.1 Participants, Apparatus, and Materials

We recruited 16 local university students as participants (8 female, 8 male) whose ages ranged from 19 to 27 (Mean = 23.50, SD = 2.12). Only one participant had no VR experience before, 10 had fewer than five experiences had used VR applications for less than five times and the rest had more than five. All the participants had normal or corrected normal vision and had watched videos with time-anchored bullet comments before, but 3 participants had not experienced 360° videos. The experimental setup and testing videos were the same as in Study 1. In addition, the participant used a controller for comment insertion.

5.2 Design and Procedure

We conducted a within-subject user study by combining two display methods *SH* and *SV* with two insertion methods *SD* and *LS*. We first let the participant read the information sheet containing the goal and process of the experiment. Then the participant was asked to sign the informed consent form, confirm they had no health problem, and fill a demographic survey.

In the experiment, the participant was trained on how to wear the headset and use the controller, and the eye tracking module was calibrated for each participant. Before the formal testing, we ensured the participant had been familiar with each of the insertion operations by taking two training trials on the same *Sheep Sports Game* video. To help the participant focus on the interaction during the training trials, existing comments for each video were hidden. Afterwards, the participant took 2 $\times$ 2 formal testing trials. In each trial, the participant was aware of the candidate comment list for the current video, and was asked to insert at least three comments to those positions where they felt appropriate in the 360° video. The order of the two chosen display methods was randomized, and the order of insertion methods was counterbalanced under each display method. The four testing videos were also randomly ordered, with each video being experienced only once. Both pre-test and post-test questionnaires were collected for each trial. A short interview was finally conducted to collect participants' subjective opinions. The whole experiment for a participant took around 26 minutes.

We recorded visual fixation and gaze movement as in Study 1. We also record the numbers of video pauses, revoked insertions (undos), and completed insertions per minute, as well as the duration of each insertion in seconds, as objective measures. For subjective measures, pre-SSQ test [17] was required before each trial [17], and post-SSQ, UEQ-S and Raw TLX [9] questionnaires were collected after each trial. After completing all the formal trials, the participant reported the favorite insertion method under each display method, and shared subjective feedback.

Table 1: Means and standard deviations on objective measures in Study 2. Code $*$ indicates significant differences between the corresponding measures, with $p < 0.05$.

Method	VF inside ROIs	VF on Comments	Gaze Movement (rad/s)	Duration (s)	Insertions (cnt/m)	Pauses (cnt/m)	Undos (cnt/m)
SH-SD	0.607±0.144	0.254±0.081	1.175±0.561	1.657±0.221	5.567±2.481	0.294±0.851	0.227±0.645
SV-SD	0.611±0.097	0.242±0.060	1.101±0.573	1.618±0.554	6.435±2.951	0.160±0.359	0.175±0.577
SH-LS	0.566±0.190	0.232±0.046	1.127±0.374	1.138±0.300	6.566±3.326	0.862±1.931	0.350±0.960
SV-LS	0.609±0.142	0.240±0.046	1.082±0.307	1.405±0.611	6.163±2.821	0.199±0.466	0.292±0.590
Overall *SD*	0.609±0.023	0.248±0.014	1.138±0.136	1.638±0.067$*$	6.001±2.717	0.227±0.646	0.201±0.603
Overall *LS*	0.587±0.035	0.236±0.012	1.104±0.074	1.272±0.105$*$	6.365±3.041	0.530±1.422	0.321±0.784

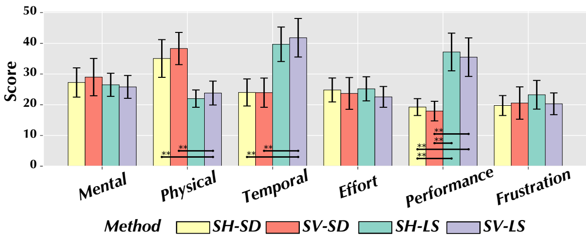


Figure 6: Means and standard errors of Raw TLX questionnaire scores on all measures in Study 2. Statistically significant differences are indicated with $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

5.3 Results

We collected data from 64 trials (16 participants × 2 display methods × 2 insertion methods). Kolmogorov-Smirnov tests revealed that the data was normally distributed for all subjective measures. Therefore, repeated measures ANOVA (RM-ANOVA) tests at the 5% significance level and Bonferroni-adjusted post-hoc tests were adopted to analyze these measures. Greenhouse-Geisser corrections were performed if the sphericity assumption was violated. For the objective data which was non-normally distributed, we used non-parametric Friedman tests to analyze the effect. If a significant effect was found, Wilcoxon signed-rank tests with Bonferrorni correction were conducted for post-hoc analysis. Table 1 illustrates the means and standard deviations of objective measures.

Visual fixation inside ROIs. No significant interaction effect of *display* × *insertion* was found ($F(1,15) = .534, p = .476, \eta^2 = .034$). Besides, *insertion* ($F(1,15) = .535, p = .476, \eta^2 = .034$) and *display* ($F(1,15) = .465, p = .506, \eta^2 = .030$) did not have a main effect on this measure.

Visual fixation on comments. There was no significant *display* × *insertion* interaction effect ($F(1,15) = .497, p = .491, \eta^2 = .032$). Neither *insertion* ($F(1,15) = .718, p = .410, \eta^2 = .046$) nor *display* ($F(1,15) = .021, p = .886, \eta^2 = .001$) has a significant main effect on this measure.

Gaze movement. Results exhibited that *display* × *insertion* interaction effect was not significant ($F(1,15) = .052, p = .823, \eta^2 = .003$). No significant main effect from *insertion* ($F(1,15) = .167, p = .688, \eta^2 = .011$) or *display* ($F(1,15) = 1.310, p = .270, \eta^2 = .080$) was observed.

Operation duration. No significant *display* × *insertion* interaction effect was found ($F(1,15) = 3.595, p = .077, \eta^2 = .193$). There was a significant main effect from *insertion* method ($F(1,15) = 11.922, p = .004, \eta^2 = .443$) on the operation duration. *LS*'s duration was significantly shorter than *SD* ($-.366, p = .004$).

Frequency of insertions. No significant interaction between *display* and *insertion* was found ($F(1,15) = 1.193, p = .292, \eta^2 = .074$). No significant main effect was revealed from *insertion* ($F(1,15) = .193, p = .667, \eta^2 = .013$) or *display* ($F(1,15) = .008, p = .931, \eta^2 = .001$).

Frequencies of video pauses and revoked insertions. The data of the two measures were not normally distributed. We thus conducted Friedman tests on combined *display-insertion* methods *SH-SD*, *SV-SD*, *SH-LS*, and *SV-LS*. No significant main effect was revealed on the frequency of video pauses ($\chi^2(3) = 1.759, p = .624$) or revoked insertions ($\chi^2(3) = .800, p = .849$). Since very few participants paused the video (12 out of 64 trials) or revoked insertion operations (11 out of 64 trails) , the standard deviations on these two measures were very large.

Table 2: UEQ-S results show the pragmatic rating, hedonic rating, and overall rating of methods in the combined display and insertion study. ">avg." represents "above average".

Method	Pragmatic	Hedonic	Overall
SH-SD	1.29 (>avg.)	1.48 (good)	1.39 (good)
SV-SD	1.38 (>avg.)	1.53 (good)	1.45 (good)
SH-LS	1.25 (>avg.)	0.86 (>avg.)	1.05 (>avg.)
SV-LS	1.39 (>avg.)	0.92 (>avg.)	1.16 (>avg.)

Workload. To analyze the data from Raw TLX questionnaires, non-parametric Friedman tests were used under all sub-measures. Significant effects were revealed for the four conditions on the physical ($\chi^2(3) = 13.684, p = .003$), temporal ($\chi^2(3) = 11.400, p = .010$) and performance ($\chi^2(3) = 15.935, p = .001$) sub-measures. Although *SV-LS* performed significantly better than *SH-SD* ($p = .008$) and *SV-SD* ($p = .001$) under physical sub-measure, it was significantly worse than *SH-SD* ($p = .007$) and *SV-SD* ($p = .005$) under temporal sub-measure. In terms of the performance sub-measure, both *SH-SD* and *SV-SD* were found significantly better than *SH-LS* ($p = .006$ for *SH-SD*, $p = .002$ for *SV-SD*) and *SV-LS* ($p = .007$ for *SH-SD*, $p = .006$ for *SV-SD*). Figure 6 shows the Raw TLX scores.

User experience. The subjective responses to UEQ-S questionnaires were collected for analysis. For the pragmatic quality, *SH-SD*, *SV-SD*, *SH-LS* and *SV-LS* were all above average, which means they had no differences in the functions of completing tasks. *SH-SD* and *SV-SD* were found benefiting the hedonic measure, while the relative metric on UEQ-S of *SH-LS* and *SV-LS* were just above average. It indicated that *SD* was more interesting and enjoyable. Furthermore, *SH-SD* and *SV-SD* had "good" overall experiences and were better than *SH-LS* and *SV-LS*, indicating that *SD* provides a better user experience in general. The results are listed in Table 2.

Simulator sickness. The SSQ data was not normally distributed (evidence from a Kolmogorov-Smirnov test). A Wilcoxon signed-rank test at the 5% significance was used to analyze the scores of pre-SSQ and post-SSQ. No significant differences were found. Nevertheless, post-SSQ scores (overall score 3.915) were all higher than pre-SSQ scores (overall score 2.805).

User Preference. 10 participants preferred *SD* under *SH* display, and the other 6 preferred *LS* under *SH*. When the display method was *SV*, 9 participants preferred *SD* and the other 7 preferred *LS*.

5.4 Interview Findings

We gathered comments from participants for further analysis, which can be summarized with the following three aspects:

Comparison between the two insertion methods. The majority of participants preferred *SD* to *LS* due to its high accuracy and naturalness: *"When using SD, I can accurately drop the comment on* *my target position even without pausing the video. (P3, P4)" "SD is more intuitive and interactive, similar to using a mouse. (P6)" "When using LS, if I do not pause the video, the inserting position may mismatch with the dynamic video content after the comment selection. (P1, P5)"* Other participants preferred *LS* because it is simple and time-saving: *"LS is more convenient and simpler, and I did not perceive any difference in accuracy between these two methods. (P10)" "Compared to LS, SD took more time for comment insertion, which can make me miss the exact timestamp. (P8)" "I prefer to locate the target position first. (P16)"*

Other related factors Only one participant proposed that different display methods affected the performance of comment insertion (*SD* outperformed *LS* under *SH* while *LS* was better than *SD* under *SV*. Most participants agreed that the display method did not affect their insertion preference. When asked about whether the experience was relevant to the video content, 10 participants answered "yes", for example: *"SD is more accurate especially when the video content changes frequently. I have to frequently pause the video when using LS. (P3)" "LS is better if the video is relatively static. (P3, 4, 7)"* Some other considerations were also reported by participants: *"I hope I can switch to no comments mode if I want to watch certain content carefully. (P11)" "I want to use the ray to choose the comment rather than using the trackpad. (P8)" "It may be more convenient to show the comment at a random position. (P15)" "My comment can overlap with others' comments, it would be better to have an automatic spatial adjustment for the comments. (P10)"*

5.5 Discussion

Although significantly shorter operation time was observed from *LS*, the majority of participants still preferred *SD* in general (10 participants preferred *SD* under *SH* and 9 participants preferred *SD* under *SV*). According to users' open comments, the dragging process from *SD* can make the comment insertion more intuitive and accurate. Therefore **H3** is rejected. Both statistic results and user interviews confirmed there was no obvious interaction effect between display methods and insertion methods, which means **H4** can be accepted. However, participants reported that the performance of insertion methods may vary on video content, where *SD* was more suitable for videos with rapidly changing content while *LS* was better for videos with relatively static content.

Regarding objective measures of visual fixation and gaze movement, there was no significant difference between the insertion methods. Interestingly, there were no significant differences between insertion methods under the measures of video pauses and revoked insertions. We noticed that either video pausing or revoking insertions were seldom used, because the participants enjoyed the continuous flow of the video experience.

No significant SSQ score differences were found before and after trials, indicating little simulator sickness can be caused by either method. UEQ-S results indicated that the two insertion methods had similar effects on user experience, while *SD* could make users more enjoyable, hence increasing the overall performance. The reason could be that participants found dragging operations more intuitive and interesting. Raw TLX results revealed that *LS* can make participants spend less effort, which was in line with the objective duration measure. However, we noticed that participants also felt more time pressure and less satisfied with the performance of *LS*. One reasonable explanation might be that sudden position change of the selected comment when releasing the touchpad could make participants confused or disoriented. This discomfort might be relieved by introducing a gentle transition animation for the position change.

Design Insights. We provide the following insights for designing insertion methods for 360° video bullet comments:

DI3: Selecting content first and then dragging it to the target position can make the operation more intuitive and accurate. Based on the previous findings, most participants preferred to select comment content first. The dragging operation can make the process more natural and accurate.

DI4: Locating first and then selecting the content can be more efficient. In our experiment, this technique requires the least operating time and effort, which can be applied when the user wants to post plenty of comments in a short time.

6 Limitations

Our work has several limitations. First, we only proposed several representative display and insertion methods for bullet comment on 360° videos. Designs related to other attributes (e.g., color, size, and transparency) are also worth studying. In addition, we assumed that bullet comments bring users a better experience. A comparison to the no-comment method could be conducted in the future. Bullet comments have different spatial linkages with the video content, whether they affect the users' preferences in display methods can be further studied. About the proposed selection-based insertion methods, although they are convenient and intuitive, users only have limited choices on the comments. In the future, feasible approaches for diversified input including virtual keyboard and speech recognition can be fully investigated. Besides, videos were played only at normal speeds in our studies. The effects of fast forward, backward and timeline switching of videos on the performance haven't been investigated. Also, there might be a potential bias due to the limited range of participants' majors and ages. In addition, only monocular 360° videos captured by still cameras were used for testing, future work could research richer types of panoramic videos, such as stereoscopic or dynamic 360° videos. Finally, video semantic understanding algorithms can be used to enhance bullet comment display and interaction. For example, object recognition and instance segmentation algorithms can be employed to avoid bullet comments obscuring visually important content of the video.

7 Conclusion

In this paper, we explored the display and insertion of bullet comments for 360° video. We designed four display methods (one planar method and three spherical methods) for VR headset and two controller-based insertion methods (*Select-and-Drag*, *Locate-and-Select*), which were evaluated via two user studies respectively. From objective and subjective measures in the first study, we conclude that spherical display methods are preferred to the planar one by offering more engagement and social interaction. Besides, spherical sliding comments are more favored than static comments because the sliding ones were more vivid, provide more audience opinions, while creating less occlusion or distraction. As revealed by our second study, the majority of participants preferred *Select-and-Drag* to *Locate-and-Select* although the latter one is less time-consuming and with less physical effort. The reason is that *Select-and-Drag* is more accurate and intuitive, making participants more enjoyable and satisfied with their operations. With these findings, we also discussed the useful insights for technique design. In the future, we would like to improve the visualization of bullet comments and investigate how to write bullet comments for 360° video in VR applications.

Acknowledgments

We sincerely thank the reviewers for their insightful comments and suggestions that helped improve the paper. This work is supported by the National Natural Science Foundation of China (Project Number: 61932003 and 61902012) and Young Elite Scientists Sponsorship Program by CAST (No. 2019QNRC001). Fang-Lue Zhang is supported by the Marsden Fund Council managed by Royal Society of New Zealand (No. MFP-20-VUW-180).

References

[1] T. Bertel, N. D. Campbell, and C. Richardt. Megaparallax: Casual 360° panoramas with motion parallax. *IEEE transactions on visualization and computer graphics*, 25(5):1828–1835, 2019.

[2] S. Cha, J. Lee, S. Jeong, Y. Kim, and J. Noh. Enhanced interactive 360° viewing via automatic guidance. *ACM Transactions on Graphics (TOG)*, 39(5):1–15, 2020.

[3] H. Chang and M. F. Cohen. Panning and zooming high-resolution panoramas in virtual reality devices. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pp. 279–288, 2017.

[4] X. Chen, Y. Zhang, Q. Ai, H. Xu, J. Yan, and Z. Qin. Personalized key frame recommendation. In *Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval*, pp. 315–324, 2017.

[5] H.-T. Cheng, C.-H. Chao, J.-D. Dong, H.-K. Wen, T.-L. Liu, and M. Sun. Cube padding for weakly-supervised saliency prediction in 360 videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1420–1429, 2018.

[6] Z. Cui, Q. Qiu, C. Yin, J. Yu, Z. Wu, and A. Deng. A barrage sentiment analysis scheme based on expression and tone. *IEEE Access*, 7:180324–180335, 2019.

[7] S. Djamasbi, A. Hall-Phillips, Z. Liu, W. Li, and J. Bian. Social viewing, bullet screen, and user experience: a first look. In *2016 49th Hawaii international conference on system sciences (HICSS)*, pp. 648–657. IEEE, 2016.

[8] X. Hao, S. Xu, and X. Zhang. Barrage participation and feedback in travel reality shows: The effects of media on destination image among generation y. *Journal of Destination Marketing & Management*, 12:27–36, 2019.

[9] S. G. Hart. Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, pp. 904–908. Sage publications Sage CA: Los Angeles, CA, 2006.

[10] M. Hosseini. View-aware tile-based adaptations in 360 virtual reality video streaming. In *2017 IEEE Virtual Reality (VR)*, pp. 423–424. IEEE, 2017.

[11] M. Hosseini and V. Swaminathan. Adaptive 360 vr video streaming: Divide and conquer. In *2016 IEEE International Symposium on Multimedia (ISM)*, pp. 107–110. IEEE, 2016.

[12] H.-N. Hu, Y.-C. Lin, M.-Y. Liu, H.-T. Cheng, Y.-J. Chang, and M. Sun. Deep 360 pilot: Learning a deep agent for piloting through 360 sports videos. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1396–1405. IEEE, 2017.

[13] J. Huang, Z. Chen, D. Ceylan, and H. Jin. 6-dof vr videos with a single 360-camera. In *2017 IEEE Virtual Reality (VR)*, pp. 37–44. IEEE, 2017.

[14] J. Jung, B. Kim, J.-Y. Lee, B. Kim, and S. Lee. Robust upright adjustment of 360 spherical panoramas. *The Visual Computer*, 33(6):737–747, 2017.

[15] R. Jung, A. S. J. Lee, A. Ashtari, and J.-C. Bazin. Deep360up: A deep learning-based approach for automatic vr image upright adjustment. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1–8. IEEE, 2019.

[16] K. Kang and S. Cho. Interactive and automatic navigation for 360 video playback. *ACM Transactions on Graphics (TOG)*, 38(4):1–11, 2019.

[17] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.

[18] J. Kopf. 360 video stabilization. *ACM Transactions on Graphics (TOG)*, 35(6):1–9, 2016.

[19] W.-S. Lai, Y. Huang, N. Joshi, C. Buehler, M.-H. Yang, and S. B. Kang. Semantic-driven generation of hyperlapse from 360 degree video. *IEEE transactions on visualization and computer graphics*, 24(9):2610–2621, 2017.

[20] G. A. Lee, T. Teo, S. Kim, and M. Billinghurst. A user study on mr remote collaboration using live 360 video. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 153–164. IEEE, 2018.

[21] Y.-C. Lee, W.-C. Lin, F.-Y. Cherng, H.-C. Wang, C.-Y. Sung, and J.-T. King. Using time-anchored peer comments to enhance social interaction in online educational videos. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 689–698, 2015.

[22] D. Li, R. Du, A. Babu, C. D. Brumar, and A. Varshney. A log-rectilinear transformation for foveated 360-degree video streaming. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2638–2647, 2021.

[23] J. Li, G. Huang, and Y. Zhou. A sentiment classification approach of sentences clustering in webcast barrages. *Journal of Information Processing Systems*, 16(3):718–732, 2020.

[24] M.-Y. Liao, C.-Y. Sung, H.-C. Wang, and W.-C. Lin. Virtual classmates: Embodying historical learners' messages as learning companions in a vr classroom through comment mapping. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 163–171. IEEE, 2019.

[25] Y.-T. Lin, Y.-C. Liao, S.-Y. Teng, Y.-J. Chung, L. Chan, and B.-Y. Chen. Outside-in: Visualizing out-of-sight regions-of-interest in a 360 video using spatial picture-in-picture previews. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pp. 255–265, 2017.

[26] B. Luo, F. Xu, C. Richardt, and J.-H. Yong. Parallax360: Stereoscopic 360 scene representation for head-motion parallax. *IEEE transactions on Visualization and Computer Graphics*, 24(4):1545–1553, 2018.

[27] G. Ma, S. Li, C. Chen, A. Hao, and H. Qin. Stage-wise salient object detection in 360° omnidirectional image via object-level semantical saliency ranking. *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3535–3545, 2020.

[28] C. Marañes, D. Gutierrez, and A. Serrano. Exploring the impact of 360 movie cuts in users' attention. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 73–82. IEEE, 2020.

[29] S. Moezzi, A. Katkere, D. Y. Kuramura, and R. Jain. Immersive video. In *Proceedings of the IEEE 1996 Virtual Reality Annual International Symposium*, pp. 17–24. IEEE, 1996.

[30] A. Nguyen, Z. Yan, and K. Nahrstedt. Your attention is unique: Detecting 360-degree video saliency in head-mounted display for head movement prediction. In *Proceedings of the 26th ACM international conference on Multimedia*, pp. 1190–1198, 2018.

[31] C. Nguyen, S. DiVerdi, A. Hertzmann, and F. Liu. Collavr: collaborative in-headset review for vr video. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pp. 267–277, 2017.

[32] H. L. O'Brien, P. Cairns, and M. Hall. A practical approach to measuring user engagement with the refined user engagement scale (ues) and new ues short form. *International Journal of Human-Computer Studies*, 112:28–39, 2018.

[33] T. Rhee, L. Petikam, B. Allen, and A. Chalmers. Mr360: Mixed reality rendering for 360 panoramic videos. *IEEE transactions on visualization and computer graphics*, 23(4):1379–1388, 2017.

[34] T. Rhee, S. Thompson, D. Medeiros, R. Dos Anjos, and A. Chalmers. Augmented virtual teleportation for high-fidelity telecollaboration. *IEEE transactions on visualization and computer graphics*, 26(5):1923–1933, 2020.

[35] J. K. Rout, K.-K. R. Choo, A. K. Dash, S. Bakshi, S. K. Jena, and K. L. Williams. A model for sentiment and emotion analysis of unstructured social media text. *Electronic Commerce Research*, 18(1):181–199, 2018.

[36] A. Schmitz, A. MacQuarrie, S. Julier, N. Binetti, and A. Steed. Directing versus attracting attention: Exploring the effectiveness of central and peripheral cues in panoramic videos. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 63–72. IEEE, 2020.

[37] A. Serrano, I. Kim, Z. Chen, S. DiVerdi, D. Gutierrez, A. Hertzmann, and B. Masia. Motion parallax for 360 rgbd video. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):1817–1827, 2019.

[38] A. Serrano, V. Sitzmann, J. Ruiz-Borau, G. Wetzstein, D. Gutierrez, and B. Masia. Movie editing and cognitive event segmentation in virtual reality video. *ACM Transactions on Graphics (TOG)*, 36(4):1–12, 2017.

[39] Y. Shen, H. Chan, and I. Hung. Let the comments fly: The effects of flying commentary presentation on consumer judgment. In *Proceedings of the 35th International Conference of Information Systems*, 2014.
[40] Y.-C. Su, D. Jayaraman, and K. Grauman. Pano2Vid: Automatic cinematography for watching 360° videos. In *ACCV*, 2016.
[41] S. Sun, F. Wang, and L. He. Movie summarization using bullet screen comments. *Multimedia Tools and Applications*, 77(7):9093–9110, 2018.
[42] A. Tang and O. Fakourfar. Watching 360° videos together. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, p. 4501–4506, 2017.
[43] A. Tang, O. Fakourfar, C. Neustaedter, and S. Bateman. Collaboration with 360° videochat: Challenges and opportunities. In *Proceedings of the 2017 Conference on Designing Interactive Systems*, p. 1327–1339, 2017.
[44] C. Tang, O. Wang, F. Liu, and P. Tan. Joint stabilization and direction of 360° videos. *ACM Transactions on Graphics (TOG)*, 38(2):1–13, 2019.
[45] J. Tarko, J. Tompkin, and C. Richardt. Real-time virtual object insertion for moving 360 videos. In *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*, pp. 1–9, 2019.
[46] J. O. Wallgrün, M. M. Bagher, P. Sajjadi, and A. Klippel. A comparison of visual attention guiding approaches for 360 image-based vr tours. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 83–91. IEEE, 2020.
[47] L. Wang, H. Wang, D. Dai, J. Leng, and X. Han. Bidirectional shadow rendering for interactive mixed 360° videos. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 170–178. IEEE Computer Society, 2021.
[48] M. Wang, Y.-J. Li, W.-X. Zhang, C. Richardt, and S.-M. Hu. Transitioning360: Content-aware nfov virtual camera paths for 360° video playback. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 185–194. IEEE, 2020.
[49] M. Wang, G.-Y. Yang, J.-K. Lin, S.-H. Zhang, A. Shamir, S.-P. Lu, and S.-M. Hu. Deep online video stabilization with multi-grid warping transformation learning. *IEEE Transactions on Image Processing*, 28(5):2283–2292, 2018.
[50] L. Xie, Z. Xu, Y. Ban, X. Zhang, and Z. Guo. 360probdash: Improving qoe of 360 video streaming using tile-based http adaptive streaming. In *Proceedings of the 25th ACM international conference on Multimedia*, pp. 315–323, 2017.
[51] S. Yamaguchi, N. Ogawa, and T. Narumi. Now i'm not afraid: Reducing fear of missing out in 360 videos on a head-mounted display using a panoramic thumbnail. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 176–183. IEEE, 2021.
[52] X. Yang, W. Binglu, H. Junjie, and L. Shuwen. Natural language processing in "bullet screen" application. In *2017 International Conference on Service Systems and Service Management*, pp. 1–6. IEEE, 2017.
[53] J. Young, T. Langlotz, M. Cook, S. Mills, and H. Regenbrecht. Immersive telepresence and remote collaboration using mobile and wearable devices. *IEEE transactions on visualization and computer graphics*, 25(5):1908–1918, 2019.
[54] Z. Zeng, C. Xue, N. Gao, L. Wang, and Z. Liu. Learning from audience intelligence: Dynamic labeled lda model for time-sync commented video tagging. In *International Conference on Neural Information Processing*, pp. 546–559. Springer, 2018.
[55] S. Zhang, H. Jiang, and J. M. Carroll. Integrating online and offline community through facebook. In *2011 International Conference on Collaboration Technologies and Systems (CTS)*, pp. 569–578. IEEE, 2011.
[56] Z. Zhang, Y. Xu, J. Yu, and S. Gao. Saliency detection in 360 videos. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 488–503, 2018.
[57] G. Zoric, L. Barkhuus, A. Engström, and E. Önnevall. Panoramic video: design challenges and implications for content interaction. In *Proceedings of the 11th european conference on Interactive TV and video*, pp. 153–162, 2013.
[58] O. Zorzi. Granovetter (1983): The strength of weak ties: A network theory revisited. In *Schlüsselwerke der Netzwerkforschung*, pp. 243–246. Springer, 2019.