Tell Me Where to Look: Investigating Ways for Assisting Focus in 360° Video

Yen-Chen Lin[12], Yung-Ju Chang[3], Hou-Ning Hu[1], Hsien-Tzu Cheng[1], Chi-Wen Huang[1], Min Sun[1]

[1]Dept. of Electrical Engineering
National Tsing Hua University

[2]Dept. of Computer Science
National Tsing Hua University

[3]Dept. of Computer Science,
National Chiao Tung University

1sunmin@ee.nthu.edu.tw 3armuro@cs.nctu.edu.tw

{[12]yenchenlin, [1]eborboihuc, [1]hsientzucheng, [1]kitsune0125}@gapp.nthu.edu.tw

ABSTRACT

360° videos give viewers a spherical view and immersive experience of surroundings. However, one challenge of watching 360° videos is continuously focusing and re-focusing intended targets. To address this challenge, we developed two Focus Assistance techniques: Auto Pilot (directly bringing viewers to the target), and Visual Guidance (indicating the direction of the target). We conducted an experiment to measure viewers' video-watching experience and discomfort using these techniques and obtained their qualitative feedback. We showed that: 1) Focus Assistance improved ease of focus. 2) Focus Assistance techniques have specificity to video content. 3) Participants' preference of and experience with Focus Assistance depended not only on individual difference but also on their goal of watching the video. 4) Factors such as view-moving-distance, salience of the intended target and guidance, and language comprehension affected participants' video-watching experience. Based on these findings, we provide design implications for better 360° video focus assistance.

Author Keywords

360-degree videos; Focus Assistance; Video Experience; Auto Pilot; Visual Guidance

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous;

INTRODUCTION

360-degree (referred to as 360°) video, also known as immersive video, is known to give viewers a spherical view and an immersive experience of the surrounding of the camera (see Figure 1 for illustration). Because of this advantage, 360° video is gaining increasing attention. Not only that video content providers have produced numerous 360° videos, online video and social media platforms such as YouTube[1] and Facebook[2] have also allowed viewers to upload and view 360° videos. However, one key issue of watching 360° videos is viewers losing track of the target to which they are intended to attend (referred to as intended target). For example, when viewers are watching a continuously and fast moving object in a 360° video, such as watching an extreme-sport video (e.g., skateboarding, rollerblading, etc.), viewers may have difficulty catching the object up and thus lose track of it. We refer to this type of challenging task as Continuous-Focus. Another challenging task is Re-Focus, where viewers need to attend to a specific location in a 360° video currently being referenced/introduced (e.g., a historical building being introduced in a tour video) while they are currently exploring other parts of the video. This task is especially challenging when the viewer need to quickly identify the location of the intended target and then attend to it when the timing is crucial (e.g., missing the introduction if the viewer is not able to attend to it in time). Failing either of these two tasks are likely to dramatically harm viewers' experience in watching 360° videos because an intended target is usually

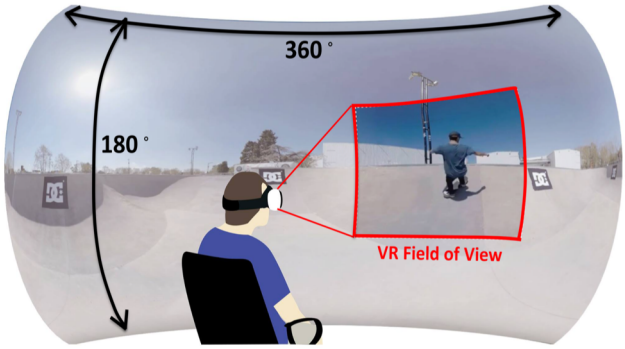

Figure 1 Illustration of the video viewing sphere (360° horizontally and 180° vertically) and the Field-of-View (red box) observed by the viewer using the VR device.

 ACM 978-1-4503-4655-9/17/05…$15.00 DOI: http://dx.doi.org/10.1145/3025453.3025757

[1] www.youtube.com

[2] www.facebook.com

the main focus of a video at a specific time. To help viewers successfully attend to an intended target, we investigated two representative Focus Assistance techniques: *Auto Pilot (AP)*—taking the viewer directly to the intended target, and *Visual Guidance (VG)*—guiding the viewer to the intended target by displaying a visual indicator signaling the direction to which the viewer should move their view. We conducted an experiment to evaluate these two Focus Assistance techniques in watching two videos—a video of an extreme sport (SPORT) primarily involving Continuous-Focus, and a video of a city tour (TOUR) primarily involving Re-Focus. We aim to answer: which technique would provide viewers with better viewing experience when focus assistance is desired.

In this paper, we report on the results of the experiment and provide key design implications for future 360° video focus assistance. Our highlights include: 1) most participants consider both AP and VG improved ease of focus in SPORT and TOUR videos. 2) AP is more suitable than VG in watching SPORT. 3) Participants have varied preferences of and experiences with AP and VG for watching TOUR, which are largely dependent on their goals of watching the tour video. 4) Participants who preferred AP highly valued being able to follow the intended focus. Participants who preferred VG highly valued having a freedom in moving their own view. 5) AP with a natural speed was more favored than with a high speed because the latter generally led to more discomfort; VG with the indicator appearing in advance was more favored than not in advance. 6) There are main effects of total view-moving distance and the speed of AP, respectively. There is also an interaction effect between maximum view-moving distance and AP speed on the viewer's discomfort. To the best of our knowledge, we are the first to investigate the effectiveness of different Focus Assistance techniques for viewing 360° videos.

The contributions of this paper are: 1) results from the first experiment investigating two Focus Assistance techniques: Auto Pilot and Visual Guidance, in assisting viewers in a Continuous-Focus type and a Re-Focus type of 360° videos. 2) Identifying specificity of Focus Assistance technique to video content with both quantitative and qualitative support. 3) Identifying the key role of the goal of watching a 360° video on viewers' choice of a Focus Assistance technique. 4) A set of design implications for assisting focus in watching 360° videos. Below, we describe the implementation of the two Focus Assistance Techniques.

TWO FOCUS ASSISTANTCE TECHNIQUES

We built an Android 360° video player using OpenGLES and implemented AP and VG on top of the player. At a high level, when playing a 360° video, we continuously tracked viewer's viewpoint position on the rendered spherical scene using a gyroscope sensor and then project the position to an equirectangular coordinate system in order to know where the viewer is looking at. Then, we use this information to calculate the shortest distance between the viewer's center view and the intended target to determine in which direction the viewer should be directed to. Below we provide more details of the two Focus Assistance techniques, along with two technique variations for each technique that we used in the experiment for comparison.

Auto Pilot (AP)

Given a known direction, AP rotated a rendered scene and directly brings the viewer to the position of the intended target when it is about to appear (see Figure 2-Left). We developed two variations for AP—AP High Speed and AP Normal Speed. We decided to test these two variations because we assumed that while AP High Speed can help the viewer quickly attend to an intended target, it may increase the viewer's discomfort. In contrast, while AP Normal Speed may introduce relatively less discomfort, it may not be able to bring the viewer to the target in time if he or she is far away from the target.

We conducted a pilot test with 10 participants to explore two rotation speeds for the experiment because we didn't find literature suggesting an appropriate speed. We let the participants experience different rotation speeds (360°, 240°, 180°, 120°, 60°/second) in a random order and chose two distinct speeds that were reported to cause less discomfort by our pilot-test participants. We finally chose 180°/second for the High Speed and 60°/second for the Normal speed.

Visual Guidance (VG)

Given a known direction, VG adds a visual indicator (see green arrow in Figure 2-Right) on the screen to guide the viewer. The visual indicator disappears when the intended target is clearly visible in the view, i.e. reaching 30° left or right to the center of field of view (FoV). Similarly, we developed two variations for VG—VG Advance Notice, and VG Normal Notice. The two variations differed in the timing for providing a visual indicator. In the pilot test, we originally let participants experience different times earlier (0.5, 1, 2, 3 seconds) to receive an advance notice in a random order. At the end, we determined that VG Normal Notice showed the indicator one second before an intended target is referenced/introduced. And VG Advance Notice showed the indicator according to when the intended target is referenced and where the viewer's current view is. The advance time was calculated using the speed of 1 second/60°, same as the AP Normal Speed. For example, if the viewer is 120° away from the intended target, VG Advance Notice shows the indicator two seconds earlier.

Instrumentation

We used a Samsung Gear VR for our smartphone mount. Viewers will need to wear the mount to see our post-processed videos. The Gear VR features precision lenses with a 96° field of view. The Gear VR weighs 310 grams. We used a Samsung Galaxy S7 smartphone with an Exynos 8890 CPU (2.3 + 1.6 GHz octa-core) and a Mali-T880 MP12 GPU which can render 3D simulations with a high

Figure 2 Illustration of two focus assistants on two types of videos: SPORT and TOUR, respectively. All red boxes denote example of a viewer's Field-of-View (FoV) in VR device. Left-panel: Auto Pilot is applied to help continuously focus on the moving intended target (skateboarder) in the Sport video. We overlaid images of three temporal locations (t_1, t_2, t_3) to highlight the motion of the intended target. Right-panel: Visual Guidance (the augmented green arrow) is applied to help refocus on the intended target described in the Tour video. We show two examples of guidance from the left and right towards the intended target in the middle. Note that the guidance dynamically changes according to the spatial relation between a viewer's head pose and the intended target.

frame rate. The Galaxy S7 features a STMicroelectronics LSM6DS3 six-axis sensor (gyro + accelerometer).

THE USER STUDY

We conducted an experiment to evaluate the variations of AP and VG in watching two 360° videos involving Continuous-Focus and Re-Focus. We also included a Baseline condition—without any technique—to examine whether any of these variations gives better viewer experience than without providing any focus assistance.

The first video was an extreme sports video (SPORT), in which participants were to continuously track a skateboarder (see Figure 2-Left). The major difficulty participants faced was Continuous-Focus. The second video was an Amsterdam city tour video (TOUR), where participants were to not only learn about several historical landmarks but also to get an immersive experience in the tour (see Figure 2-Right). The tour video contained 14 intended target occasions, i.e., introductions of specific places, locations, and objects, with the rest being periods during which viewers can freely browse the surrounding. The major difficulty participants faced was Re-Focus—attending to an intended target after changing scenes and during free browsing, respectively. It is noteworthy that unlike in SPORT, where the intended target was mainly the skateboarder, in TOUR, we had to determine the intended target based on the narrative script of the video tour. We defined that an object was an intended target when the script used the term "see," "look at," "this is", "here is" and so on to refer to the object or indicated the specific location of the object (e.g., "on the right side is the place XYZ").

Our evaluation measures were primarily related to video-viewing experience and discomfort because we thought these are what an ideal Focus Assistance technique needs to optimize. Our measures included ease of focus, engagement, enjoyment, feeling of presence, and discomfort. Below, we describe the experiment design.

Experiment Design

We divided the experiment into two sessions (shown in Figure 3), one session watching SPORT, and the other watching TOUR. The order of the two sessions was randomized and counterbalanced. We separated SPORT and TOUR because we did not compare Focus Assistance variations for SPORT. It was because in this video the viewer is continuously tracking the skateboarder, where the attention is presumably not too far away from the skateboarder. Thus, differentiating speeds in AP and timings of notice in VG would only make a slight difference. As the result, in the SPORT session, participants watched the video three times (see Figure 4 (a)): Baseline, AP, and VG, with the order of the latter two randomized. (we did not include Baseline in the randomized order because we focused on the comparison between AP and VG. Making participants watch the Baseline first was to let them more easily felt the assistance of the four techniques later. The SPORT video clip was 55 seconds long and watched without sound.

In the TOUR session, we evaluated two variations of AP and VG, respectively. Thus, there were in total four technique variations being compared in TOUR. Each participant was assigned a randomized order of the four techniques: 1) AP High Speed, 2) AP Normal Speed, 3) VG Advance Notice, and 4) VG Normal Notice. We split the TOUR video into four clips, and each clip was watched with sound using one of these technique variations. We processed and cut the four clips in a way that each clip had a similar characteristic: they were all of similar length (55, 56, 53, 75 seconds respectively); contained at least one long period where they could explore the surroundings; and contained at least each of the two types of Re-Focus occasions—changing scenes and referencing object. Note that we could not make clips in an identical length for the experiment because we directly used videos available online. Since we did not want to modify the original content

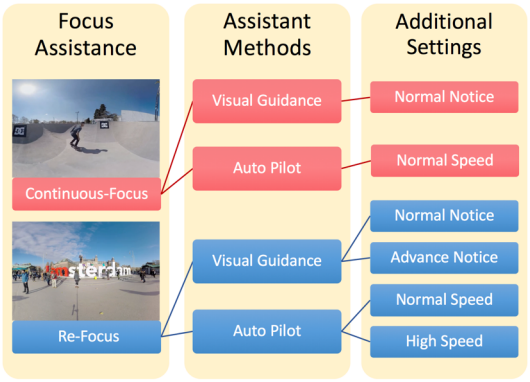

Figure 3 Experiment Setting. The experiment was in two sessions: watching Continuous-Focus video (i.e. SPORT), and watching Re-Focus video (i.e. TOUR). The order of the two sessions were randomized and counterbalanced among participants. In SPORT, participants were exposed to the normal version of Auto Pilot (AP) and of Visual Guidance (VG), respectively. In TOUR, participants were exposed to in total four technique variations of AP and VG. They are: AP High Speed, AP Normal Speed, VG Advance Notice, and VG Normal Notice. The order among these four was randomized and counterbalanced. Each of these technique was contrasted to a Baseline in watching the same video clip.

of the video, the actual length of each clip largely depended on the storyline of the video and on the appropriateness of a breakpoint between two clips. We chose not to split the video into four identical-length video clips because doing so would make the storyline disconnected and the breakpoints interruptive, which then harmed the viewer experience. Under this constraint, we tried our best to make the mentioned characteristic of the clips as similar as possible.

Similarly, we let participants watch the Baseline before they watched a clip with assistance to help them more easily feel the assistance. As a result, participants watched eight clips in the TOUR session. In the entire experiment, each participant watched in total eleven clips. Note that we chose not to let participants watched the entire TOUR video for each technique because we assumed they would be less motivated to explore in the tour after they had seen the video several times. This might affect their enjoyment and engagement of watching the video. In addition, watching the entire TOUR video for each technique would make the video watching unnecessarily long, letting participants likely to be more uncomfortable after the study.

For each of these eleven sessions, participants filled a video viewing experience questionnaire where they assessed ease of focus, engagement, enjoyment, feeling of presence, and discomfort related measures (general discomfort, feeling of vomit, feeling of dizziness) (see Figure 4 (b)). The inclusion of these measures was inspired by previous research and the Simulator Sickness Questionnaire (SSQ) [6,8] (we chose not to use the entire SSQ because our participants needed to fill a questionnaire in total 11 times.

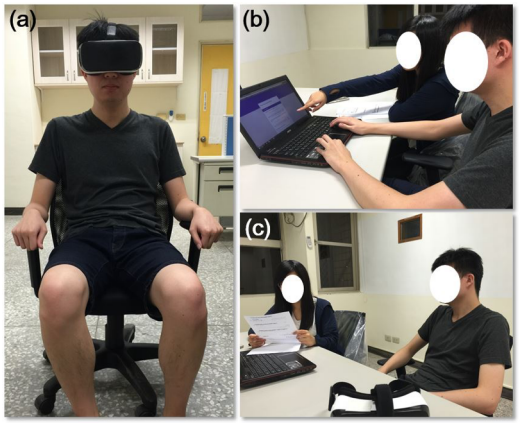

Figure 4 Illustration of our environment setting and experiment procedure. For each session, participants first (a) watched videos with a VR device on a swivel chair. Then (b) participants filled in a questionnaire. After participants going through one of SPORT or TOUR video, we (c) conducted a debriefing interview about their experiences.

Filling out the entire SSQ would be a great burden for them). After participants completed watching SPORT and TOUR, we conducted a debriefing interview to ask their qualitative experiences (see Figure 4 (c)). We first asked them to explain and analyze what they had observed in the clips to check whether they had noticed the differences in clips caused by the assistance. Then we asked how they felt about the assistance from each technique variation; in what aspect they liked or disliked the technique; and their ranking of the techniques. Finally, we asked their feelings about which technique was more suitable for which video.

Participants

We recruited 32 participants via a subject pool (16 females). 30 participants were from the universities of the co-authors. All of them were between 19 and 34 years old (M=21.93, SD=2.68). Eleven of them have experience in using a VR device for a particular purpose (10 on watching a video, 9 on playing games). However, all participants have quite limited experience in using a VR device (rating themselves as using it less than once a month).

QUANTITATIVE ANALYSIS AND RESULTS

We analyzed participants' self-assessed experiences using a mixed effect ordinal logistic regression. We chose this technique because of a within-subject design and that all the measures were ordinal variables. We included the-order-of-the-video variable to account for the order effect. We added a random factor of viewers to account for individual differences. Below we report the qualitative results.

Ease of Focus

Both AP and VG were rated to have better ease of focus than Baseline in SPORT and in TOUR (*SPORT*: AP vs. Baseline: $Z(13)=5.19$, $p<.001$; VG vs. Baseline: $Z(13)=2.42$, $p=.02$; AP (M=8.19, SD=2.25), VG (M=6.28, SD=2.76), Baseline (M=5.31, SD=2.32). *TOUR*: AP vs. Baseline: $Z(13)=6.27$, $p<.001$; VG vs. Baseline:

Z(13)=6.77, p<.001; AP (M=7.61, SD=2.29), VG (M=7.63, SD=2.30), Baseline (M=4.96, SD=2.56)). In SPORT, AP was also rated to achieve better ease of focus than VG (Z(13)=4.19, p<.001). In TOUR, there was no significant difference among the four technique variations.

Engagement

Participants felt less engaged with VG than with Baseline and with AP when watching SPORT. (VG vs. Baseline: Z(13)=-2.27, p=.02; VG vs. AP: Z(13)=3.80, p<.001, AP (M=7.41, SD=1.83), VG (M=6.09, SD=2.05), Baseline (M=7.03, SD=2.29)). AP did not differ from Baseline in this aspect (Z(13)=0.70, p=.49). In TOUR, however, both AP and VG made participants feel significantly more engaged than Baseline (AP vs. Baseline: Z(13)=4.61, p<.001; VG vs. Baseline: Z(13)=4.14, p<.001; AP (M=7.16, SD=1.90), VG (M=7.03, SD=1.77), Baseline (M=5.91, SD=1.94)). There was no significant difference among the four technique variations.

Enjoyment, Presence, and Receptivity

Participants had better enjoyment, feeling of presence, and receptivity to video content in SPORT with AP than with VG and Baseline, respectively (*Enjoyment*: AP vs. Baseline: Z(13)=4.70, p<.001; AP vs. VG: Z(13)=4.10, p<.001; AP (M=6.97, SD=2.19), VG (M=5.31, SD=2.01), Baseline (M=5.03, SD=2.29). *Feeling of Presence*: AP vs. Baseline: Z(13)=2.87, p=.004; AP vs. VG: Z(13)=3.42, p<.001; AP (M=6.38, SD=2.28), VG (M=5.13, SD=2.25), Baseline (M=5.31, SD=2.52). *Receptivity*: AP vs. Baseline: Z(13)=3.56, p<.001; AP vs. VG: Z(13)=4.13, p<.001; AP (M=7.19, SD=1.89), VG (M=5.63, SD=1.79), Baseline (M=5.75, SD=1.80)). There was no significant difference between VG and Baseline in these aspects in SPORT. In TOUR, participants had better enjoyment and receptivity to video content with both AP and VG than with Baseline, but not feeling of presence (*Enjoyment*: AP vs. Baseline: Z(13)=2.45, p=.01; VG vs. Baseline: Z(13)=969.09, p=.006; AP (M=6.53, SD=2.09), VG (M=6.66, SD=1.87), Baseline (M=5.86, SD=1.94). *Receptivity*: AP vs. Baseline: Z(13)=4.81, p<.001; VG vs. Baseline: Z(12)=-4.63, p<.001; AP (M=7.03, SD=1.91), VG (M=7.08, SD=1.55), Baseline (M=5.84, SD=1.76)). We did not see a significant difference in enjoyment, feeling of presence, and receptivity in TOUR among four techniques variations.

Discomfort

Participants rated significantly less discomfort with AP than with Baseline and VG in SPORT (AP vs. Baseline: Z(13)=-4.20, p<.001; AP vs. VG: Z(13)=-3.02, p=.003; AP (M=2.84, SD=2.65), VG (M=3.81, SD=2.83), Baseline (M=4.28, SD=2.87)). VG only had a marginal effect (Z(13)=-1.89, p=.06). In TOUR, on the contrary, participants rated more discomfort with AP than with VG (Z(13)=2.12. p=.03, AP (M=3.27, SD=2.71), VG (M=2.75, SD=2.48), Baseline (M=3.04, SD=2.55)). Note that we did not analyze feeling of vomit and dizziness because these items are two of the many items used for measuring overall simulator sickness [6]. Since we did not include other items from SSQ, we ultimately only analyzed self-rated general discomfort. We further looked into the impact of the distance participants' view traveled in watching the video and its interaction effect with the AP speed. Our results did not show any main effect of average, median, and the maximum of distance on discomfort. However, we found a main effect of total distance, a main effect of speed, and an interaction effect between maximum distance and AP speed. More specifically, the farther participants' view moved in total, the more discomfort they felt (Z(18)=2.10. p=.04). Participants also felt less discomfort using AP Normal Speed than AP High Speed (Z(18)=-2.47. p=.01). However, participants felt more discomfort when the maximum distance was moved with Normal Speed (Z(18)=2.66. p=.007).

To summarize, the results showed that using either AP or VG, participants felt it easier to focus in SPORT and in TOUR than without assistance (Baseline). In comparing AP and VG, AP performed better than VG for Continuous-Focus. Second, both AP and VG made participants felt more engaged, enjoying, and receptive to video content than Baseline in TOUR. However, in SPORT, only AP made participants felt more engaged than Baseline. When using VG, participants even felt less engaged than Baseline. Third, both AP and VG improved participants' enjoyment and receptivity to video content in TOUR; however, they did not improve their feeling of presence. Fourth, AP made participants feel less discomfort in SPORT but more discomfort in TOUR. We also found a main effect of total distance, AP speed, and an interaction effect between AP speed and maximum distance. Finally, we did not see any significant difference in all aspects among the four technique variations for TOUR videos.

In conclusion, these quantitative results seem to suggest the specificity between technique and video content. That is, for Continuous-Focus video (e.g., SPORT), AP performed better than VG almost in all aspects. For Re-Focus video (e.g., TOUR), in contrast, these results suggested no obvious advantage of any technique variation of AP and VG. Our qualitative findings below provide some insights.

QUALITATIVE ANALYSIS AND RESULTS

We transcribed the audio-recorded debriefing interviews and two researchers coded the transcriptions using an iterative process of generating, refining, and probing emergent themes. The coding achieved 0.87 inter-rater reliability using Cohen's Kappa, indicating very good agreement between the two coders[7]. Because participants had quite different reactions and responses to the techniques used for between SPORT and TOUR, we discussed their feedback regarding these two separately.

Auto Pilot and Visual Guidance for the SPORT Video

While the quantitative results suggested that AP outperformed VG in most aspects for the SPORT video, participants' qualitative experience further explained why such an advantage of AP over VG existed. Specifically, 30

out of 32 participants reported that they preferred AP over VG for watching SPORT. Common reasons for preferring AP included that it allowed them to track the skateboarder at all times and it reduced their effort. For example, U10 said, *"The video would move with the skateboarder. You don't really need to move much."* U28 also said, *"you don't need to keep moving, which makes you more concentrate on the content."* U31 added, *"Auto view is awesome. You totally don't need to move much but you can see clearly. So you can better enjoy and know what the video is doing."* In addition, according to many participants, AP also reduced discomfort, as U19 said, *"Helping you turn is much better. I already felt dizzy when I wore the VR device. I felt even more if I had to look for the person. [With Auto Pilot] you don't need to turn your head much, and you don't need to look for the person."* U17 also said, "*The skateboarder video, I felt auto rotation helped quite a lot. I felt dizzy when I rotated by myself."* Interestingly, when being asked whether they had noticed a difference between Baseline and the video where AP was used, a few participants did not know why it was easier to track the skateboarder, for example, U15 stated, *"The third one [Auto Pilot] was like magic. You could just catch up. I have no idea why!"*

On the contrary, most participants thought VG provided limited help. For example, U10 commented, *"The skateboarder was too fast. Even if you gave me the hint, I still couldn't get it."*. U17 commented, *"The arrow [indicator] was not helpful. Of course, you knew it's moving which way. The point is that you still couldn't catch up."* U21 even thought the guidance was not helpful at all, *"The arrow made me feel that I was cheated. I turned there, then it told me to turn to the other side. And then I saw nothing."*

In addition, a few participants thought the indicator was a distraction when they were tracking the skateboarder. For example, U16 said, *"With the cue [indicator], you have to keep watching when and where it appears. You don't need to do this for the Auto Pilot."* U15 said, *"I think the arrow is somehow annoying because it keeps blocking the scene. I'm already trying to get that person, but it's still there."* The experienced distraction might explain why participants felt less engaged in SPORT than AP and even Baseline.

To summarize, participants' qualitative experience showed a clear preference to AP over VG. And this result is consistent with the quantitative results that AP outperformed VG in all aspects except engagement. However, participants' qualitative experiences uncovered reasons that caused these differences.

Auto Pilot and Visual Guidance for the TOUR Video

Unlike the dominant preference to AP over VG for SPORT, participants had more balanced preferences between AP and VG for the TOUR video. The balanced preferences can be observed from participants' ranking of the four technique variations. First of all, while 13 participants ranked AP Normal Speed as their top one, 12 participants ranked VG Advance Notice as their top one. Second, 15 participants' top two choices included one variation of AP and one variation of VG, instead of placing two variations of either AP or VG as their top two. Third, 11 participants preferred VG over AP (i.e. placing both VG variations as their top two), and 6 ranked the opposite. These observations suggested that participants have more balanced preferences between VG and AP. To look further, out of those who placed both AP and VG in their top two, participants thought both AP and VG helped them focus on the intended target and receive and comprehend the video content better, despite the difference in how they were assisted. For example, U16 said, *"sometimes I couldn't hear the guide well and you didn't know what she wanted me to look at. With [both] assistance you would realize 'oh! This is what you're saying.' It helped a lot."* U11 said, *"auto rotation just made things in front of you when the guide was talking. Well, both [AP and VG] made it easier. You could also follow the arrow to find things easily."*

However, other participants had a stronger preference toward either AP or VG. Participants who preferred AP mostly favored following the guide of the tour. For example, U30 said, *"You were brought to see the building first, and then you listened. This was more helpful."* U17 explained why she preferred following the guide, *"It is a tour. I'd think all it wants you to know is the thing the guide introduces. If you pass those things, the guide is meaningless."* The most commonly cited advantage of AP was its fast and precise focus, which helped participants better comprehend the guide and concentrate on the content. U19 said, *"You jump to the guide so that you immediately know what she's talking about."* U18 said, *"I felt auto rotation gave better concentration, because it immediately brought you to the point, and you knew that's what you should be paying attention to."* The high precision of focus was considered especially useful when the intended target was among similar adjacent objects, as U9 highlighted, *"Auto focus is better because it just brings you to the thing. The arrow (VG) lets you know the thing is within a range, but you didn't know which building."*

Participants who preferred VG generally disliked AP for its lack of freedom, which they considered important for a city tour. For example, U16 said, *"I think freedom is very important for a tour. […] I don't enjoy being forced to see what I don't wanna see. You can just tell me what is there. But I don't need to look at them now."* U16 continued, *"the arrow gives you more freedom. You can choose not to turn and look at what interests you."* U6 explained why it was important to have the freedom, *"Because it's a tour. You'd like to see the whole Amsterdam instead of just what she is introducing."* U21 also said, *"The auto rotation didn't give you the feeling of looking around. You had to follow its rhythm. I prefer to look around by myself."*

Other people preferred VG because AP let them have a feeling of sudden. U22 said, *"It moved really fast. I was*

like shocked when my view got suddenly moved." VG, according to participants, also made them more prepared. U27 said, *"You know the indicator, and then you see the thing. It makes you more prepared."* U1 also said, *"It's like a left turn light before you left turn. You gotta let people prepare for that."*

Within AP, more (23) participants liked AP Normal Speed than AP High Speed. The major reason mentioned was that High Speed made them felt uncomfortable, as U27 explained *"The fast one made me very dizzy. […] it is more helpful, but at that moment [of moving] I felt very uncomfortable. The slow speed at least had some transition. And during that, you still knew what the guide was talking about."* However, participants who preferred AP High Speed thought it was more efficient than the Normal Speed, as U5 said *"The slow speed was way too slow. I didn't even know whether it's actually moving."*

On the other hand, more (19) participants preferred VG Advance Notice than VG Normal Notice. The most common cited reason was having more time for moving their view, as U25 reported, *"If the arrow appears in advance, I have more time to turn my view."* U2 also explained, *"If you see the notice earlier, you can move your view. After 1-2 seconds, you are at the moment. If you get the notice at the time the guide talks, you probably cannot catch the point."* In contrast, participants who preferred VG Normal Notice desired to know the content at the moment instead of earlier. For example, U20 said, *"If the cue comes too early, I don't know what to look at. […] I am an impatient person."* U21 also said, *"I'm afraid that if the clue shows too early, there would be no surprise."* Some other participants thought whether showing the notice in advance depended on the target. For example, U14 said, *"For things always there, like building, it [advance or not] doesn't make a difference to look at it two seconds earlier or later. But if you're looking at something that would move, showing the arrow earlier would be better."*

Finally, there were also comments regarding other factors affecting participants' viewing experience and preference. For instance, despite appreciating the usefulness of easier focus, a number of participants thought both assistance techniques were unrealistic and unnatural. For example, U7 commented on the AP in SPORT, *"I felt being able to catch up the skateboarder is kind of unreal."* U12 commented on AP in TOUR, *"Taking me to see the other part makes me feel like your world is controlled by other people."* U28 also commented on VG in TOUR, *"I think putting the arrow there makes the scene less harmonious."* We think these reactions might explain why participants did not get a better feeling of presence with the assistance in watching TOUR. In addition, a few participants thought the visual indicator made them feel being passive, U31 said, *"The arrow made me become more passive. I'd wait until the arrow came out, and did not think how I should turn by myself."* Another factor mentioned by several participants was the comprehension of video content. For example, U4 suspected that his language comprehension might have been affecting his preference, "*After some thought, I think I probably would have preferred the arrow if I had better understood the English [the guide is speaking]."*

To summarize, we see a number of agreements between qualitative and quantitative results. For example: AP was more favored than VG for SPORT, and participants preferred AP with a slower speed. In addition, the varied preferences of and experiences with AP and VG in TOUR might explain the absence of statistically significant difference but discomfort among the four technique variations in TOUR. We found that, in particular, the varied preferences and experiences seemed to be linked not only to individual differences but also to the participants' goal of watching the video, such as to precisely and quickly follow the guide or the narrative of the video, or to freely explore the environment without any enforcement.

DISCUSSION

Our results offered both quantitative and qualitative support that in general, providing viewers with focus assistance, helps both Continuous-Focus and Re-Focus in 360° videos. In addition, both AP and VG made participants feel more engaged, enjoying, and receptive to the video content in TOUR. However, there were factors we found influential on participants' preference of using AP and VG in a 360° video. We discuss these factors in the following sections.

The Video Content Matters

First of all, our results highlighted the specificity of Focus Assistant techniques to video content. That is, when an intended target is a single, continuous, and fast moving object, AP displayed a great advantage in improving the ease of focus over VG. Moreover, AP also increased participants' engagement and reduced their discomfort compared to Baseline while VG failed at both. Such specificity is supported by not only the quantitative results but also participants' qualitative experience. According to the participants, the observed advantage of AP over VG in SPORT was mainly because they only needed to focus on one instead of multiple targets in the video. Since the intended target always remained present and the same in SPORT, AP made participants' field of view nearly "attached" to the skateboarder at all time. This largely reduced their effort of moving their body or head to catch up the skateboarder. In contrast, because VG provided only a visual indicator, participants needed to mainly rely on themselves to continuously track the skateboarder. As one participant put, *"seeing it being there is one thing; being able to catch it up is another thing."*

In TOUR, however, it took participants less effort to attend to an intended target because most of the intended targets in TOUR were static. In addition, the fact that participants had to switch among multiple intended targets instead of one and that participants only needed to occasionally, instead of continuously, attend to an intended target largely reduced

AP's advantage. In addition, probably because participants had to "jump" among different objects several times with AP, participants rated higher discomfort with AP than with VG. As a result, compared to watching SPORT, VG was preferred by more participants in TOUR. Consequently, we think our data indicated a crucial role of video content and highlighted its specificity to Focus Assistance .

The Goal of Watching the Video Matters

Our data also highlighted an influence of the goal of watching the video on participants' preference of and experience with a Focus Assistance technique. For example, in SPORT, while most participants appreciated being able to continuously track the skateboarder easily, a few participants felt that such an ease was unrealistic. One participant mentioned that she preferred to pursue the skateboarder by herself without any assistance.

The impact of goal was especially obvious in TOUR. While some participants thought it was important to follow the tour guide during the tour, others preferred to spend more time looking around and exploring the environment on their own. As mentioned earlier, participants of the former preferred AP because it helped them quickly and precisely attend to the intended target being introduced by the guide. Participants of the latter, however, highly valued the freedom for self-exploration and disliked being "forced" to change their current view. Some of these participants reported that they were not always listening to a guide when they were traveling because the content was not always interesting to them. These two distinct goals of watching a video tour might explain why we did not see any statistically significant difference among the four technique variations in all aspects but discomfort in TOUR.

Other Considerations for Choosing Focus Assistance

In addition to the two highlights of the specificity and the goal of watching the video, our data also suggested a number of factors to take into consideration when choosing a type of focus assistance. The first factor is the distance between the viewer's current view and the intended target. Our results suggested that the total distance the participant's view has moved was positively correlated with discomfort. In addition, participants' discomfort was also positively correlated to the interaction between the largest distance the view traveled and AP Normal Speed. In other words, when the viewer is far away from the intended target (e.g., >90°), using AP with a slower speed means that the viewer has to travel for longer time, which thus aggregated more discomfort. However, because the relationship between distance and speed could be complex, future work is needed to explore the influence of distance and speed further.

The second factor was the salience of the intended target. Several participants mentioned a challenge of identifying the intended target among similar adjacent objects. This challenge had made many participants preferred AP over VG in TOUR because AP brought participants directly and precisely to the intended target and saved their time looking for the intended target on their own. Making the indicator directly pointing to the target might be a solution.

The third factor was the salience of the visual indicator in contrast to the video background when VG was used. For example, according to our observation and to the qualitative data, participants would pay attention to the visual indicator when it appeared. As a result, when the indicator was not apparent enough viewers may have trouble with noticing it. On the other hand, some participants mentioned that the visual indicator was annoying and distracting, and would block the view when they had already known where to move their view to. As a result, future work is to explore a good timing for automatically hiding the visual indicator. However, we think a simple way is allowing the viewer control the visibility of the visual indicator.

Fourth, we found language comprehension to be another factor to consider. In our study, the language of TOUR was English while all participants were non-native English speakers. Although participants did not need to understand all the vocabularies the tour guide said and that most participants self-reported that they had recognized the keywords and knew what was being referenced, we believed language ability might still have an impact on participants' engagement and receptivity to the video content, as well as their preference of the techniques, despite the fact that the within-subject design might have reduced this individual difference. We argue that the language factor needs to be acknowledged and considered for choosing a type of focus assistance because it is common, and soon will be more common, for a viewer to watch a video in their non-native language (e.g. English). For example, one participant suspected that he might have liked VG better if he had better understood the guide. On the other hand, if the viewer cannot comprehend the video at all, it is also likely that they prefer VG and explore in the video tour all by themselves. We think future work is needed to further understand the impact of language comprehension on the effectiveness of focus assistance.

Finally, although our data did not show any main effect of the video order on the self-rated measures, we think the length of time, i.e., how long participants have worn a VR device to watch a 360° video might have an impact on their experience and discomfort. In our study, we asked participants to take breaks in order to reduce the effect of time. However, in real life settings, viewers would watch videos at their own pace. Unfortunately, we did not examine the effect of the length of time.

In general, our study suggests that in SPORT, viewers In the section below, we consider the factors aforementioned and offer design implications for future focus assistance technique to support watching 360° videos.

Design Implications

At a high level, we suggest developing a hybrid Focus Assistant combining AP and VG that allows viewers to

switch on and off certain features such as auto rotation and visual indicator. The Assistant should also allow viewers to adjust parameters such as speed, salience of the visual indicator, timing of showing the indicator, and so on. We suggest full user control because viewers' goal of watching a video is situated and is only known by viewers themselves. Because their goals may change their preference, granting them the control would be necessary. However, to reduce viewers' burden, we suggest the Assistant choose a default technique based on the type of content. For example, during a period of Continuous-Focus, such as tracking a single and continuously moving target, we suggest using AP as a default because our data showed that AP outperformed VG in many aspects and it led to less discomfort. However, for videos like TOUR, we suggest video providers offer information of intended targets so that the Assistant knows when an intended target would appear and when assistance should be provided. For videos without such information, the Assistant can infer the intended target by analyzing the narrative (i.e., subtitle) of the video. When the video approaches the time when an intended target will appear, the Assistant shows the direction of the intended target with a salient visual indicator in advance and starts to detect the movement of the viewer's view. If the viewer shows no intention to move, the Assistant by default removes the indicator and waits for the next intended target; however, if the viewer is detected to be moving toward the intended target, the Assistant assists moving with a natural speed by default and allows speed adjustment by the viewer at any time. The purpose of this is to show a visual indicator in advance so that the viewer can prepare for the move. When the intended target has entered the viewer's field of view, the Assistant highlights the intended target to help the viewer know its exact location. The Assistant removes the highlight after a certain amount of time so that the visual indicator will not block the scene. We believe granting user control and detecting the movement of the viewer's view can address the goal of watching a video. We believe such a hybrid Assistant can improve viewers' experience in watching a Continuous-Focus as well as a Re-Focus type of 360° videos while not harming their freedom to explore the environment. It is worthwhile to examine the actual effectiveness of this proposed Assistant on a prototype in future research.

STUDY LIMITATIONS

The current study is subject to a number of limitations. First, this work concentrated on videos with intended target(s). There are 360° videos purposed mainly for sharing the immersive experience without needing the viewer to attend to any specific intended target (e.g., capturing the experience of downhill mountain bike riding, roller coaster riding). In these videos, viewers may not consider AP and VG as useful they are for the SPORT and TOUR videos. Second, we also did not examine the two Focus Assistance techniques in videos involving multiple intended targets concurrently shown in the video. Third, we only measured participants' experience in two 360° videos, and there was only one type of focus task used in each. In other words, we did not examine videos involving both Continuous-Focus and Re-Focus, or even other focus tasks, in one video. We also chose only one particular video for each. Fourth, we only examined two assistance techniques and two variations of each. It is possible that we could have observed more or fewer differences had we tested more variations. Fifth, the experiment was conducted in a setting where participants watched the video with a VR headset. Our results regarding discomfort thus might not apply to watching 360° videos without wearing a VR headset. Sixth, all participants who watched the English videos were not English native speakers. If we had let the participants watch videos in their own native language, the results could have been different. Finally, we did not quantitatively examine the influence of factors such as length of time, language comprehension, salience of indicator, familiarity of subject and salience of the intended focus on video watching experience. Instead, we identified these factors based on participants' qualitative feedback and on our observation of them watching the videos. Nevertheless, we argue that it is worth further investigating in future research to better understand how to choose a suitable focus assistance technique.

RELATED WORK

As mentioned earlier, to the best of our knowledge, we are the first to investigate the effectiveness of different Focus Assistance techniques for improving the viewer's experience in watching a 360° video involving an intended target. As 360° videos have not been emergent until recently, it is not surprising that only a limited number of approaches have been introduced and developed for this purpose. We review related work in this research space.

360° videos are a new medium for telling a story interactively. As Vosmeer and Schouten [11] suggested, unlike two distinct engagement styles—lean-back while watching a movie vs. lean-forward while playing a video game, 360° videos enable video providers to design an intended engagement at every moment in a video. Inspired by Vosmeer and Schouten [11], Gugenheimer et al. [3] proposed a motorized swivel that allowed viewers to fully explore the surrounding (lean-forward engagement) in viewing 360° videos. At some specific moments, the motorized chair could also nudge viewers' orientation (lean-back engagement) to specific events or scenes. Note that this work differed from the product called RotoVR [12] in that RotoVR is a motion platform controlled by a gamepad for game playing. Recently, Facebook has introduced Guide [13], which is intended to let content providers of 360° videos set a narrative for their videos by highlighting specific points of interest over the course of the videos. Once a set narrative is used, viewers are automatically directed around the video as it plays.

Thus, both Gugenheimer et al. [3] and Facebook Guide share the same idea with our Auto Pilot in directing viewers

to an intended target. However, whereas Gugenheimer et al. [3] argued that rotating a virtual scene in front of the viewers' eyes will lead to simulator sickness [3], our results suggested that although Auto Pilot indeed caused more discomfort in the Re-Focus type of video (TOUR), it reduced participants' discomfort in the Continuous-Focus type of videos. We think the difference may be that while Gugenheimer et al. [4] rotated the "real body" of the viewer, our Auto Pilot rotates the view *per se*. Furthermore, our findings and design implications are useful for Facebook Guide to better serve a vast number of 360° video viewers.

At the other end, an article by *5 Lessons Learned While Making Lost* [14] by Oculus [15] suggested that one should neither control nor guide the attention of viewers. Instead, they should let the viewers discover the story by themselves. Nevertheless, our study results offer another perspective: Whether or not it is worth providing a focus assistance depends on the viewers' personal preference, the content of the video, and the goal of watching the video. As we have provided abundant evidence for the benefit of providing focus assistance in the context of watching SPORT and TOUR videos where the viewers are intended to attend to a specific target at a specific moment, we argue that the guideline by [14] may not be suited to the context of this particular study. Moreover, according to what we have found in the study—the important role of the specificity and the goal of watching the video, we argue against a general guideline and argue for leaving viewers the control over which focus assistance to use and when to use them.

Finally, Sheikh et al. [9] conducted a study to evaluate the effectiveness of different techniques for directing viewers' attention in 360° videos. However, instead of designing post-processing techniques (e.g., Auto Pilot and Visual Guidance), they evaluated the techniques at filming videos such as the motion, gestural or audio cues of a bystander.

It should be noted that outside of the 360° videos world, other researchers have attempted to solve an off-screen visualization problem on mobile devices. Our Visual Guidance technique has been inspired by these works such as Overview+Detail [1] and Contextual Cues [2]. An Overview+Detail [1] visualization displays an overview simultaneously with a detailed view in a separate window. Usually, the overview shows the entire space at the reduced scale and includes a properly positioned graphical highlight to indicate the portion of space currently shown in the detail view. For example, in the Large Focus-Display [5], the overview is a miniature version of the information space that uses a rectangular viewfinder to highlight the currently displayed portion.

Unlike Overview + Detail visualizations which needs multiple windows, Contextual Cues visualizations focus on providing appropriate information to locate intended targets even when they are off-screen. These approaches display abstract shapes (or proxies) in a border region of the screen to serve as visual references to intended targets outside the viewer's view area. For example, Burigat et al. [2] uses scaled and stretched arrows that encoded distance information of off-screen intended targets as size and length of arrows. Gustafson et al. proposed Wedge [4], which conveys direction and distance information of off-screen intended targets uses acute isosceles triangles in order to avoid overlap and clutter.

Despite their similarity to Visual Guidance, our study looked at 360° videos with only one intended target rather than multiple off-screen objects. In addition, in supporting watching 360° videos, it is crucial to keep viewers comfortable when choosing a focus assistance technique. 360° video is unique in that viewers would need to rotate their head to focus on an intended target rather than simply sliding the touch screen.

Finally, researchers such as Song et al. [10] have attempted to develop techniques (e.g., zooming and enhancing regions of interest) to enhance the video-watching experience on mobile devices. However, this line of works focused on standard-view videos instead of 360° videos, which therefore do not have the challenge of focusing tasks that our study attempted to address. We exclude them from the literature review because of the different focus.

CONCLUSION

We present Auto Pilot and Visual Guidance and evaluated their effectiveness in supporting watching an extreme-sport (SPORT) and a video tour (TOUR), that involved two types of focus challenge—Continuous-Focus and Re-Focus, respectively. Our results showed that with either AP or VG, participants felt it easier to focus on the intended target in watching both videos. Furthermore, our results highlighted two important factors influencing participants' preference of and experience with a Focus Assistance technique.

The first is the specificity of Focus Assistance technique to video content. While Auto-Pilot outperformed Visual Guidance almost in all aspects when watching SPORT, it is less advantageous in watching TOUR. Second, we highlighted the role of the goal of watching a video in affecting participants' preferences of and experience with focus assistance. Participants who valued following the narrative of a tour preferred Auto Pilot for its precise and fast focus. Participants who valued exploration on their own preferred Visual Guidance for having more freedom. Based on these findings, we provide design implications for future focus assistance to improve the viewing experience of 360° videos.

ACKNOWELEDGEMENT

We thank Novatek Microelectronics Corp. for their support. We also thank research collaborators including Ming-Yu Liu and other HCI peers in Taiwan for their feedback. We thank our reviewers for providing constructive suggestions.

REFERENCES

1. Stefano Burigat and Luca Chittaro. 2011. Visualizing references to off-screen content on mobile devices: A comparison of Arrows, Wedge, and Overview + Detail. *Interacting with Computers* 23, 2: 156–166.
2. Stefano Burigat, Luca Chittaro, and Silvia Gabrielli. 2006. Visualizing Locations of Off-screen Objects on Mobile Devices: A Comparative Evaluation of Three Approaches. In *Proceedings of the 8th Conference on Human-computer Interaction with Mobile Devices and Services*, 239–246.
3. Jan Gugenheimer, Dennis Wolf, Gabriel Haas, Sebastian Krebs, and Enrico Rukzio. 2016. SwiVRChair: A Motorized Swivel Chair to Nudge Users' Orientation for 360 Degree Storytelling in Virtual Reality. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 1996–2000.
4. Sean Gustafson, Patrick Baudisch, Carl Gutwin, and Pourang Irani. 2008. Wedge: Clutter-free Visualization of Off-screen Locations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 787–796.
5. B. Karstens, R. Rosenbaum, and H. Schumann. 2005. Presenting large and complex information sets on mobile handhelds. *E-commerce and M-commerce Technologies*: 32–56.
6. Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. 1993. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology* 3, 3: 203–220.
7. Klaus Krippendorff. 2004. *Content analysis: An introduction to its methodology*. Sage.
8. J. J. W. Lin, H. B. L. Duh, D. E. Parker, H. Abi-Rached, and T. A. Furness. 2002. Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. In *Proceedings IEEE Virtual Reality 2002*, 164–171.
9. Alia Sheikh, Andy Brown, Zillah Watson, and Michael Evans. 2016. Directing attention in 360-degree video.
10. Wei Song, Dian W. Tjondronegoro, Shu-Hsien Wang, and Michael J. Docherty. 2010. Impact of Zooming and Enhancing Region of Interests for Optimizing User Experience on Mobile Sports Video. In *Proceedings of the 18th ACM International Conference on Multimedia*, 321–330.
11. Mirjam Vosmeer and Ben Schouten. 2014. Interactive Cinema: Engagement and Interaction. In *Interactive Storytelling*, Alex Mitchell, Clara Fernández-Vara and David Thue (eds.). Springer International Publishing, 140–147.
12. Roto VR chair - interactive virtual reality seat. *Roto VR chair - interactive virtual reality seat*. http://www.rotovr.com/
13. New Publisher Tools for 360 Video | Facebook Media. https://media.fb.com/2016/08/10/new-publisher-tools-for-360-video/
14. 5 Lessons Learned While Making Lost. https://storystudio.oculus.com/en-us/blog/5-lessons-learned-while-making-lost/
15. Oculus. *Oculus*. https://www.oculus.com/