Depth Conflict Reduction for Stereo VR Video Interfaces

Cuong Nguyen[1] **Stephen DiVerdi**[2] **Aaron Hertzmann**[2] **Feng Liu**[1]

[1]Portland State University
Portland, OR, USA
{cuong3, fliu}@pdx.edu

[2]Adobe Research
San Francisco, CA, USA
{diverdi, hertzmann}@adobe.com

ABSTRACT

Applications for viewing and editing 360° video often render user interface (UI) elements on top of the video. For stereoscopic video, in which the perceived depth varies over the image, the perceived depth of the video can conflict with that of the UI elements, creating discomfort and making it hard to shift focus. To address this problem, we explore two new techniques that adjust the UI rendering based on the video content. The first technique dynamically adjusts the perceived depth of the UI to avoid depth conflict, and the second blurs the video in a halo around the UI. We conduct a user study to assess the effectiveness of these techniques in two stereoscopic VR video tasks: video watching with subtitles, and video search.

ACM Classification Keywords

H.5.1 Information Interfaces and Presentation: Multimedia Information Systems

Author Keywords

Virtual Reality; stereoscopic; 360; subtitles; video interface.

INTRODUCTION

The recent consumer availability of virtual reality (VR) head-mounted displays (HMDs) has created considerable interest in capturing and sharing 360° video, particularly stereoscopic 360° video. These videos are creating new media for entertainment [29], news and documentaries, real estate, data visualization [8], virtual tours [14], and free-viewpoint video [20]. Stereoscopic 360° video, which provides a much greater sense of immersion than monoscopic video, is becoming increasingly available due to many recent advances in camera technology [2, 30], and new processing tools [22] and editing interfaces [32, 33] are being developed for these videos.

However, the extra sense of depth in stereoscopic VR video can be problematic to users of VR video applications. Common user interface (UI) widgets like video navigation, subtitles, annotations, and tool palettes are often rendered on top of the video. Each widget must be rendered at a specific perceived depth, which is controlled by varying the disparity (difference in horizontal position) between the left and right eye views. However, the perceived depth of the stereoscopic

CHI 2018, April 21–26, 2018, Montreal, QC, Canada

 ISBN 978-1-4503-5620-6/18/04…$15.00

DOI: https://doi.org/10.1145/3173574.3173638

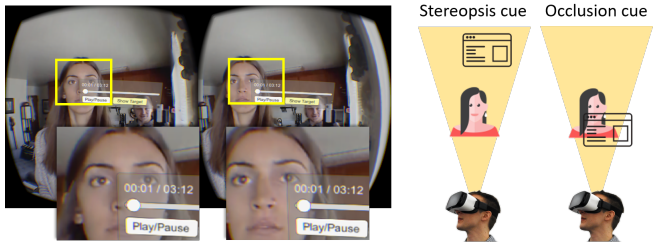

Figure 1: Depth conflict illustration. An interface (the video player) overlays a video object (the actress) but is actually behind her in depth. Left: The resulting graphics are uncomfortable to view in VR (e.g., the areas behind the text 00:01 in the insets are different between the left and right views). Viewers may also experience difficulty changing focus between the interface and the video. Right: illustration of the conflicting depth cues perceived by the same viewer in the same view. © Kevin Kunze

VR video varies greatly, which can create a conflict in perceived depth. Specifically, when a UI element is rendered over a video element that is perceived to be closer than the UI element (Figure 1), there is a conflict between the stereopsis depth cue and the occlusion depth cue. In other words, objects in the video are blocked by a widget which is *behind* the video objects. This creates visual discomfort [25, 44] and confusing visual cues. Also, alternating eye focus between the video and the UI can be difficult when the difference in depths is large. A naive approach to these problems is to place the UI very close to the viewer. However, prolonged exposure to close objects in VR is uncomfortable [23, 27] and still does not solve the problem completely: one cannot move the UI close enough to never conflict with video objects.

A key insight of our work is that, because conflicts occur at the intersection of the UI elements and the video, they can be resolved by local adjustments: either by adjusting the UI, or adjusting the video locally around the UI.

Based on this insight, we introduce two techniques for reducing these depth conflicts. The first, Dynamic Depth, dynamically adjusts the depth of UI widgets so that they normally rest at a comfortable default depth, but move closer to the viewer (by changing disparity) when video elements would conflict. The apparent depth of video elements is precomputed by a stereo computer vision algorithm.We also introduce a simpler technique, Halo Blur. Halo Blur simply blurs the video around the UI. While this does not necessarily produce a geometrically-valid configuration, it is very simple to implement and can mask high-frequency depth cues from stereo

images [13, 18], thus potentially reducing depth conflicts and helping ease focus on the interface. Our techniques are local, fast, and do not estimate eye gaze, and therefore could be used on low-end HMDs.

We evaluate our techniques in a preliminary within-subject user study with two video tasks: watching videos with subtitles and video searching. We examine performance data and subjective ratings of the participants to assess how depth conflicts affected their experience. We find that Dynamic Depth is preferred over the baseline condition and over Halo Blur.

RELATED WORK

Visual discomfort induced by stereoscopic displays in VR HMDs is an active research area. Most work has focused on the vergence-accommodation conflict (VAC). There are established solutions to this problem such as rendering virtual content within a comfortable parallax zone or using novel hardware displays. For a detailed discussion, please refer to recent surveys [23, 19]. Besides vergence and accommodation cues, the human visual system can perceive depth from other cues such as occlusion, texture, or motion. Even when VAC is resolved, depth cues can conflict and cause discomfort [27]. Our work focuses on addressing a few specific depth conflicts that arise in VR video interfaces.

Depth conflict problems have been investigated in other stereoscopic media. In film production, the *window violation* problem occurs when a video object is seen as in front of the screen but is clipped by the edge of the screen [45]. Window violation is not a problem in VR HMDs [16]. Subtitles that are placed on top of stereoscopic video can also cause depth conflicts [7, 26]. Video editors can place subtitles closer to the viewer during production, but subtitles are still problematic in live-broadcast stereoscopic TV [40]. Using subtitles in VR video poses similar problems because the perceived depth can vary greatly. In stereoscopic 3D applications, depth conflicts affect small on-screen widgets such as mouse cursors [37] or gaming crosshairs [38]. Game designers can utilize known depth information from the 3D scene to minimize conflicts [38, 39]. In contrast, our work focuses on handling depth conflicts in a dynamic VR environment, where the arrangement between the UI and the video is unpredictable.

Commercial video viewers for stereoscopic VR video do not handle depth conflicts directly. The GoPro VR Player does not display UI elements during playback. The Oculus Video application renders the video image as monoscopic when UI elements are shown on top. JauntVR's player renders timeline widgets and subtitles close to the viewer.

Dynamic stereo adjustment methods help reduce visual discomfort in stereoscopic scenes. The idea was first proposed by Ware et al. [44]. These methods adjust the focus plane such that it aligns with the object the viewer is looking at, which can help reduce VAC or enhance depth perception in 3D scenes [24, 34]. In the same spirit, in Augmented Reality research, Bell et al. first introduced view management techniques that can place textual labels at depths to avoid occlusion [4]. Our Dynamic Depth also adjusts the depth of the UI widgets dynamically. The novelty lies in tuning the adjustment based on the perceived depth of the video to reduce depth conflicts with the video content.

DEPTH CONFLICTS IN VR VIDEO INTERFACES

In this section, we identify three sources of perceptual conflict that can happen when UIs overlap stereoscopic VR video. Current VR HMDs typically use stereoscopic display to convey depth perception [23]. These displays render slightly different views to the left and right eyes, allowing the viewer to see depth through the *stereopsis* process [19]. The human visual system also perceives depth through other cues such as occlusion, motion, texture, or perspective. Viewers may experience discomfort and perceptual difficulties when different depth cues are perceived simultaneously. We use the term "conflict" loosely to mean incongruent depth cues.

Occlusion/stereopsis conflict. UI widgets are normally rendered on top of video, indicating that the UI is in front of the video. The UI and stereoscopic video are each rendered with disparities that creates perceived depth. If the UI disparity indicates that it is further away than the video elements, then the cues conflict because it is physically impossible for an object to be occluded by another object that is behind it (Figure 1). Since occlusion is one of the strongest depth cues [27], viewing video objects that are closer in depth but behind the UI can cause discomfort and some forms of double vision, e.g., the viewer sees two images of either the UI or the video [25, 31].

Near/far conflict. Some applications, like video editing, require frequent shifts of attention between the UI and the video. When the UI and the video are perceived at very different depths, the viewer's eyes have to re-verge when transitioning from one element to the other. Alternating eye vergence can be cumbersome and may cause eyestrain [26, 41].

Pictorial conflict. Some textures of the video around or behind the interface can be seen very differently between the left and right eyes (Figure 1 Left). The resulting graphics can create *binocular rivalry* [45]. Although binocular rivalry is a natural phenomenon, it is more prevalent in consumer stereo displays because they cannot render natural depth-of-field blur [18]. This phenomenon can make it difficult for the viewer to fuse the stereo images of the video, or properly focus on the interface. Moreover, many VR applications typically render the UI semi-transparent to prevent blocking too much of the content underneath. Semi-transparent UI can potentially worsen the effect of binocular rivalry [28].

TECHNIQUES CONSIDERED

We built a generic 360° stereoscopic video application environment in VR to experiment with our techniques. To enable stereoscopic playback, the left and right images of each frame are texture mapped on two 3D spheres, each of which is rendered to the left and right views of the VR HMD, respectively. The camera baseline is set to the viewer's interpupillary distance (IPD). UI elements are rendered on top of the video. We consider the UI as a 2D WIMP-style panel (Windows Icons Menus Pointers) because it is most relevant to video applications. A panel can contain various widgets such as texts, buttons or sliders. The UI can be rendered either in

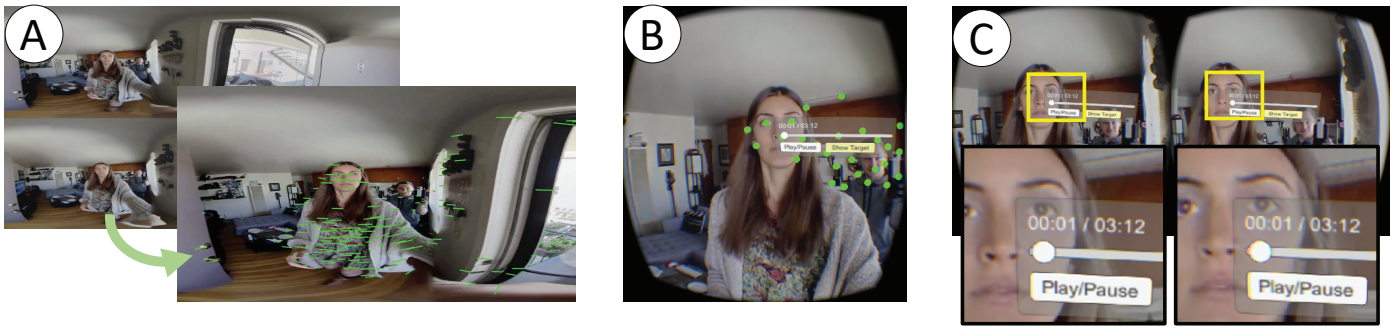

Figure 2: Overview of Dynamic Depth. (A) We pre-process the input video to find feature points and left/right disparities (e.g., green lines on the actress) (B) Features points are mapped to the VR view and shown as the green dots (only for illustrative purposes). Dynamic Depth estimates the perceived depth of the video based on these points. It detects when depth conflict occurs by comparing the depths between the UI and the video. (C) Dynamic Depth moves the UI closer to the viewer to reduce depth conflicts. Notice in the insets that the areas around the interface's corner are more geometrically consistent compared to the same scene in Figure 1. © Kevin Kunze

display-fixed (e.g., advertisement banners, compass, subtitles) or *world-fixed* mode (e.g., tool palettes, annotations, gaze triggers) [15]. Our video application is developed in Unity3D for the Oculus Rift CV1 HMD. Users use a 6DOF controller to interact with the interface.

Baseline

The baseline technique we consider is to place the UI at a fixed depth of 3 meters from the viewer, as suggested by the Oculus design guidelines [1]. Placing the UI even closer than that could further reduce depth conflicts, but would also increase oculomotor discomfort [27].

Dynamic Depth

At a high level, Dynamic Depth detects and resolves depth conflicts by adjusting the perceived depth of the UI so that it appears at the same depth of the nearby video content (Figure 2). This requires first estimating the perceived depth of the video. We also limited the rate at which the rendered depth of the UI changes, in order to avoid distractingly-fast changes.

Disparity map pre-computation

We first pre-process the input video to estimate dense correspondences between the left and right view in each frame using an optical flow method [35]. The resulting flow vectors give the disparities of each pixel in the video frame. A set of left/right feature points and their corresponding disparities are selected for the real-time conflict detection step (Figure 2A).

Potential conflict detection

Based on the disparity maps, Dynamic Depth detects depth conflict by comparing the perceived depth of the video and the UI elements in VR.

We approximate the perceived depth of the video by analyzing the region of the disparity map that the UI currently overlays. Feature points from the disaprity map are converted to spherical coordinates in the video spheres, so we can compare the position of the UI in VR to any of these points. Then, we select a subset of points within 30° from the UI's center (Figure 2B). Each point maps to a video feature around the UI and also contains disparity information. Following the approach of Blum et al. [6], the screen disparity d of a feature point is:

$$d_i = (p_{\text{focus},L} - p_{\text{focus},R}) - (p_{i,L} - p_{i,R}) \quad (1)$$

where p_{focus} is the focus point of the VR HMD (e.g., the 3D point where the disparity is zero) and p_i is the feature point that the viewer is looking at. L and R denotes the screen coordinates of these points in the left and right cameras. Smaller negative d indicates video objects that seem closer to the viewer, while positive d indicates objects that are in focus or seem farther away.

Thus, at any moment in time, we can compute the perceived depth of the video region around the UI (denoted as d_{Video}) by aggregating d values of the nearby feature points. To account for variations in depth, we extract the top 10% smallest d values and find the median d value. The data is then filtered with a moving window from the last 10 data points to reduce noise. We focus on small d values because closer video objects can attract more attention [43]. They also take up more space in the view and are more likely to cause conflicts. The final smoothed disparity value (d_{Video}) gives a proxy to estimate the perceived depth of the video around the UI.

To determine when depth conflicts occur, we also compute the screen disparity of the UI's center point (d_{UI}) using Equation 1. We consider a conflict to occur when the absolute difference between d_{Video} and d_{UI} is larger than a threshold $t = 5 pixels$. This value is chosen empirically so that Dynamic Depth can quickly trigger the next depth adjustment step. We used an absolute difference to handle both cases when the perceived depth of the UI is behind or in front of the video.

UI depth adjustment

If there is no depth conflict, the UI is placed at the default depth. Otherwise, Dynamic Depth adjusts the UI's depth to minimize the difference between d_{Video} and d_{UI}, which in turn can resolve the conflicts. However, changing the UI's 3D coordinates would require additionally adjusting scaling, collision detection, and rendering order parameters. Thus, we followed the approach of Oskam et al. and change d_{UI} by

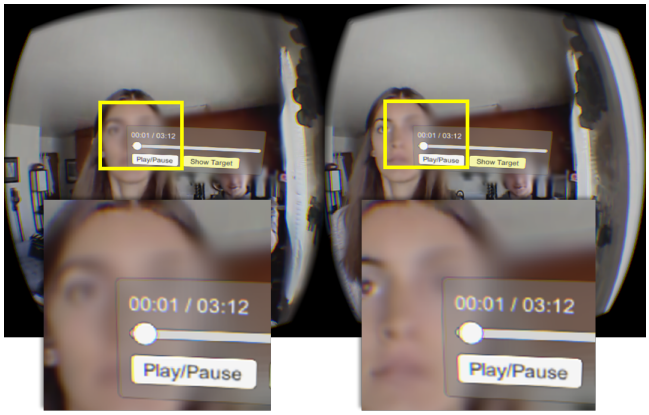

Figure 3: Halo Blur blurs the video content around the UI. Insets: compared to the same scene in Figure 1, the current scene is still not geometrically consistent. However, the blur effects mask high-frequency spatial information in the video and make the details from the UI clearer. © Kevin Kunze

shifting the left and right camera frustums horizontally [34]. Briefly, shifting the frustums inward makes the UI appear closer and vice versa (Figure 2C). These cameras are used to render the UI independently from the video, so the disparity shift does not affect the depth perception of the video.

An important consideration of camera shifting is the rate of change of the UI's depth (denoted as δ, measured in arcmin/s). Shifting the frustums too quickly can be distracting and uncomfortable [42], while shifting too slowly will make the UI unable to resolve depth conflicts in a timely manner. To balance this tradeoff, we adjust δ using a number of heuristics determined from a pilot study with four test users.

Default rate. First, we set the default δ values to be 60 arcmin/s when the UI needs to move closer to the viewer and 30 arcmin/s when it recedes. They were chosen so that the UI would move closer in depth faster when it conflicts with video elements. Otherwise, the UI does not need to act fast when there is no depth conflict or when video elements are far away, and so it can recede more slowly.

Speed up. Second, we speed up these rates by a factor of 10 when the UI's position changes quickly. Users often move the UI, either through head motion or through the hand controller. When the position changes quickly, such as when the user looks around for inspection, it is highly unlikely that the user would notice any changes in the UI because of *change blindness* phenomenon [5]. We leverage these opportunities to quickly adjust the UI depth without sacrificing comfort. We set the speed threshold to be 7 m/s based on the pilot tests.

Limits. Finally, we limit the UI depth near the comfort zone suggested by the HMD manufacture guideline [1] to avoid causing excessive uncomfortable disparities.

Halo Blur

We also explore an alternative, simpler technique that does not require the estimation of the video disparity map. Our Halo Blur technique applies blur effects to the video images around

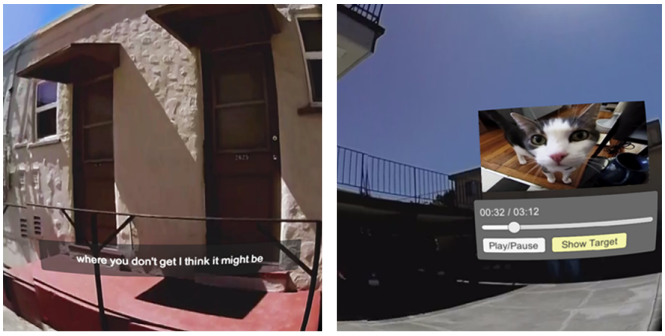

Figure 4: Illustration of the Subtitles task (Left) and the Search task (Right). Please refer to the video demo for a better assessment. © Kevin Kunze

the UI (Figure 3). The blur effect masks high-frequency spatial data from images and thus could potentially weaken the stereopsis cue or reduce binocular rivalry [13, 18]. Thus, Halo Blur could help reduce depth conflicts and ease the eye transition between the UI and the video. Compared to Dynamic Depth, it also does not require estimation of depth maps or shifting the UI in distracting ways.

To achieve the blur effect, we add an additional blurred texture canvas below the UI pane. The canvas is slightly bigger than the UI. For each pixel in this canvas, a fragment shader looks up the color values of the video image textures from the previous rendering passes. The shader then applies a Gaussian blur kernel to each pixel, varying the strength of the effect so that it is strongest in the center and gradually diminishes outward.

USER STUDY

We conducted a preliminary user study to evaluate our two techniques in two test VR video applications. Participants were asked to perform two common video tasks: watching video with subtitles (*Subtitles*) and searching for a target video scene in VR (*Search*). These tasks allowed us to examine our techniques in different application scenarios.

We performed our experiments on a Windows computer with an Intel Core i7 3.4 GHz CPU, 16 GB of memory, and an NVidia GeForce GTX 970. Participants were seated in a swivel chair. We used an Oculus Rift CV1 and a 6DOF Touch controller for the study.

Tasks

In the Subtitles task (Figure 4 Left), participants were asked to watch videos with subtitles. We lowered the video volume to 5%, so this task would demand high attention focus to follow the video's narration or dialogue. Participants were also explicitly instructed to watch the video and read the subtitles. The UI was rendered as a text box that follows the user's head movement (e.g., display-fixed) and was semi-transparent. Participants could adjust the vertical placement of the text box using a thumbstick controller.

In the Search task (Figure 4 Right), participants were asked to search for a target scene as quickly and accurately as possible. The target scene is shown as a thumbnail on the UI. We

chose the Search task in addition to Subtitles because this is a common subtask in analyzing videos (e.g., for criticism, for home movies) and, crucially, in video editing (e.g., reviewing and finding specific clips). To search, the user needs to deliberately alternate attention between the UI and the video (e.g., scrubbing in a timeline to control the search, adjusting playback to evaluate a specific clip). Many other editing tasks using similar behaviors (e.g., color grading and compositing); we chose Search as the simplest example of more active video interactions. The UI is a basic video player timeline interface, including a seek slider, a play/pause button, and another button that shows or hides the target thumbnail. The UI was rendered in world-fixed mode and was fully opaque. Participants could use a controller to move the UI or interact with the buttons.

To reduce learning effects in the study, we used six different videos. The details of these videos are summarized in Table 1. Five of them captured scenes using static cameras. One video (*Hemophillia*) contains a 3-second segment with extreme camera motions. We dimmed this segment to prevent motion sickness. These videos were selected because they contain many scenes with multiple objects that appear close to the viewer and thus can induce depth conflicts.

Experiment design & procedure

We recruited 12 participants for the study (10 males and 2 females ranging in age from 19 to 26 with a mean age 23.3) from a university. Four participants had not experienced VR before. The rest reported limited experience. We checked to make sure participants could see stereoscopic 3D with a few test scenes in the HMD. Three participants wore glasses during the study. We calibrated participant's IPD using the HMD's built-in tool (mean = 64.58 mm).

We chose a within-subject study design. Our independent variable was Technique (Dynamic Depth, Halo Blur, and Baseline). With two tasks and three techniques, each participant performed 6 trials (2 tasks $\times$ 3 techniques). We used a different video for each trial to reduce learning effects. We selected 6 videos for 6 trials. We fixed the order of the video (Table 1) but counter-balanced the order of the techniques. After each trial, participants filled out a questionnaire form with questions about symptoms of depth conflicts and their subjective preference (Table 2). We also recorded task time and task error in the Search task. The timer starts when a participant selects the "show target" button, and ends when the show target button was selected again. The Task error is the absolute difference between the target time and the participant's time (in seconds). We also measured Simulation Sickness Questionnaire (SSQ) scores [21] throughout the study to observe participants' comfort level.

The study was conducted in a university lab. Participants were first explained the procedure. Then, they performed the Subtitles and the Search task in order. Before each task, participants practiced the task using a test video for 5 minutes and filled a pre-SSQ form. In each task, the order of techniques was counterbalanced using a Latin square design. After each trial, participants filled out post-questionnaires and SSQ forms. To reduce carry-over effect, they were allowed to rest for 5 to 10 minutes between tasks. The study lasts about 45 minutes.

Results

All participants completed the study without any noticeable signs of motion sickness. In each task, each participant filled the SSQ questionnaire before the task and after each trial, resulting in 8 data points. They are: before task 1 ($M = 0.91, SD = 1.08$), after trial 1 ($M = 1.08, SD = 0.99$), after trial 2 ($M = 1.08, SD = 1.31$), after trial 3 ($M = 3.0, SD = 4.49$), before task 2 ($M = 0.5, SD = 0.79$), after trial 4 ($M = 0.33, SD = 0.65$), after trial 5 ($M = 0.33, SD = 0.65$), after trial 6 ($M = 0.66, SD = 0.98$). The ratings after each trial rise slightly compared to the rating before the task. We analyzed the differences using a paired-samples t-test and found that none of them was statistically significant ($p > 0.05$).

We then analyzed the subjective questionnaires to examine if participants experienced symptoms of depth conflicts. We used Friendman's test. Post-hoc analysis was done using Wilconxon signed rank tests. We applied Bonferroni correction to adjust for multiple testings and used an alpha level of 0.05. Figure 5 summarizes the ratings in both tasks.

In the Subtitles task, participants experienced various problems from depth conflicts. Participants gave Dynamic Depth better ratings in all questions. There were significant differences for all except Q1. We report the results below:

Q1 (Focus switch) ($\chi^2(2) = 5.31, p > 0.05$). Participants who used Baseline and Halo Blur reported having difficulties changing eye focus, especially when the pace of the subtitles is faster in the Hemophilia video. We observed that 3 participants used an ad-hoc solution to reconcile the problem: they quickly find and turn to a farther scene elsewhere in order to make the switch easier. Thus, they did not rate the problem negatively.

Q2 (Legibility) ($\chi^2(2) = 15.59, p < 0.01$). Participants who used Baseline and Halo Blur reported that they were not able to read the text when the actors or the blood cells in the video are nearby. Two participants mentioned that if they squinted their eyes then they could read it, but it was very cumbersome. In Dynamic Depth, none of the participants reported any problems. Only one participant reported that the subtitle "seems a bit too close", but it was still easy to read. The differences between Dynamic Depth and Baseline ($Z = -3, p < 0.05$) and Halo Blur ($Z = 2.94, p < 0.05$) were statistically significant.

Q3 (Double vision) ($\chi^2(2) = 15.95, p < 0.01$). When asked to describe why the subtitles were difficult to read, most participants who used Baseline and Halo Blur reported they saw two overlapping images of the subtitles. The double images made it very difficult to focus on the text. Double images did not occur in the Dynamic Depth condition. The differences between Dynamic Depth and Baseline ($Z = -2.6, p < 0.05$) and Halo Blur ($Z = 3.08, p < 0.05$) were statistically significant.

Q4 (Distraction) ($\chi^2(2) = 17.7, p < 0.01$). Participants who used Baseline and Halo Blur found that it was quite distracting when they could not read the subtitles. In Dynamic Depth, the ratings for this question are not as high. The differences between Dynamic Depth and Baseline ($Z = -2.85, p < 0.05$) and Halo Blur ($Z = 3.07, p < 0.05$) were statistically significant.

Video	Task	Description
Party1 [12]	Subtitles	Actors perform a scene around the camera and speak to the viewer. (2:05 min.)
Party2 [12]	Subtitles	The second part of Party1. (2:05 min)
Hemophilia [9]	Subtitles	A narrator explains Hemophilia as the camera moves in a 3D scene of human blood veins (2:21 min)
Gladiator [17]	Search	Two gladiators fight around the camera. The target scene is set at the beginning of the fight. (2:05 min)
Proposal [10]	Search	Two main actors and a few film crews surround the camera. The target scene is set at a unique pose of the two actors. (1:26 min)
Circus [11]	Search	Six artists perform around the camera. The target scene is set at a unique pose of three artists. (2:34 min)

Table 1: Brief descriptions of the video materials used in the study.

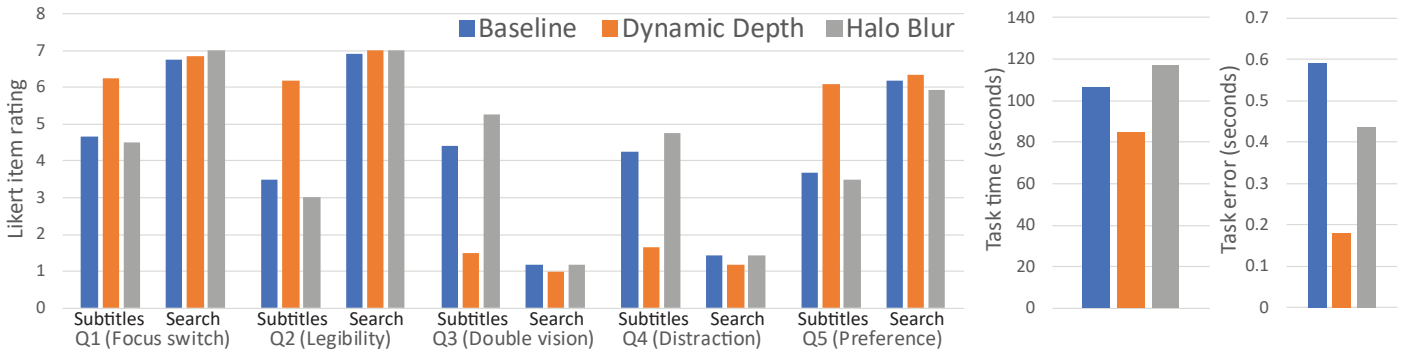

Figure 5: (Left) Summary of participants' ratings to the subjective questionnaire in both tasks. (Middle and Right) Task time and task error summary of the Search task.

Subjective Questionnaire	
Q1 (Focus switch)	How easy was it to change focus between the UI and the video?
Q2 (Legibility)	How easy was it to view the information on the UI?
Q3 (Double vision)	To what extent did you notice double images of the same object?
Q4 (Distraction)	To what extent did you think the UI was distracting during the task?
Q5 (Preference)	To what extent did you prefer this condition?

Table 2: Subjective questionnaire. Participants responded with a 7-point Likert item, ranging from very difficult to very easy (Q1,Q2) and not at all to very much (Q3,Q4,Q5).

Q5 (Preference) ($\chi^2(2) = 16.5, p < 0.01$). Finally, all participants gave higher preference ratings to Dynamic Depth. The differences between Dynamic Depth and Baseline ($Z = 2.95, p < 0.05$) and Halo Blur ($Z = -2.95, p < 0.05$) were statistically significant.

In the Search task, the differences in all questions were not statistically significant. None of the participants reported any visual problems or artifacts related to depth conflicts. We further looked into the performance data of the Search task (Figure 5). The task time and task error data were analyzed using repeated-measures ANOVA. Overall, participants in all conditions were accurate in finding the target scenes. The differences in task error were not statistically significant ($F(2,22) = 0.49, p > 0.05$). Participants who used Dynamic Depth were slightly faster than the other two conditions, but the differences were also not statistically significant ($F(2,22) = 3.14, p > 0.05$).

DISCUSSION & LIMITATIONS

In Baseline, participants' ratings about depth conflicts were different between the two tasks. In the Subtitles task, participants reported many problems that prevented them to read or focus on the texts. The UI in this task was semi-transparent, which can create strong conflicting depth cues between the text box and the video [28]. The difficulty is reinforced by the task characteristic. To read the whole sentence, participants need to focus longer on the UI, so they are more likely to see depth conflicts. In contrast, in the Search task participants did not experience depth conflicts, even though the three target scenes in this task contain large regions of close video objects. The Search task is characterized by rapid attention switching between the video and the UI, so participants might spent less time focusing on UI elements. Furthermore, the UI is opaque, so depth conflicts only occur around the edges and corners of the UI. As a results, the visual conflicts might have occurred peripherally.

These results suggest that depth conflict problems might depend on the nature of the task and the UI design. Depth conflicts are problematic when the UI is semi-transparent and when the task highly demands the viewer's attention. Depth conflicts are not as strong when the UI is opaque, or when the task can be done with peripheral vision.

Halo Blur did not work as well as expected. In the Subtitles task, even though the blur effects made the text box less transparent, we found that participants were still affected by conflicting depth cues from the video. This result is particularly interesting because blur has been known to reduce binocular rivalry [18] or weaken disparity depth cues [13]. However, most of these studies were done on static images. The text box used in our study follows the user's head movements. Thus, one potential explanation is that the users might have perceived motion cues from the video even in the blur regions. These cues could amplified the spatial information from the video images [36] and make blur less effective. We report the negative results of Halo Blur for the sake of completeness. The design and outcomes of this technique could also be interesting for future work in this area.

Dynamic Depth is a promising solution to depth conflicts. The ratings from the Subtitles task show that participants who used it reported the least problems. In the Search task, participants in both Baseline and Halo Blur conditions were slightly slower than in Dynamic Depth. We suspect that Dynamic Depth helped participants switch focus between the video and the UI faster because the UI was at the same depth of the video. Switching focus between large distances in depth often takes time [3]. However, since the current task duration is quite short, and we did not measure precise eye fixations, we could not confirm this observation.

Dynamic Depth detects depth conflicts in a conservative way by aggregating video disparity values around the UI. On one hand, it works well for small UI elements that are meant to be used together with video viewing. On the other hand, it may not work as well with more complex UI designs such as those that cover a large area of the video. The UI may overlay multiple video objects with varying depths. Integrating eye tracking to obtain more accurate video depth cues and developing solutions for large UI designs is an interesting direction and is left for future work.

Finally, we consider a few limitations of the user study. First, we did not measure participants' comprehension level in the Subtitles task. Methods to measure comprehension typically require participants to perform quizzes, which may add unnecessary stress to an already overwhelming VR experience (most of our participants had little or no experience with VR). Second, participants performed only short tasks in the study. Thus, we could not examine long-term effects of depth conflicts in VR video interfaces. This is an important research direction and should be investigated more in future. Third, the number of participants is rather low, which could limit the generalizability of our findings because of individual differences. Nevertheless, our study is a first step in exploring this problem and our results suggest a few promising insights that could be helpful for both application designers and researchers.

CONCLUSION

We explore depth conflicts between UI and stereoscopic video and discuss how they can affect user experience in VR video interfaces. We present two techniques to address this problem. Dynamic Depth detects and reduces depth conflicts by analyzing the video content and adjusting the depth of UI widgets to the depth of the video. Halo Blur simply blurs the video around the UI. We evaluated these techniques in a preliminary user study with two video tasks: watching video with subtitles and video searching. Our study compares our techniques with a baseline condition where the UI is fixed at a comfortable distance in VR. Our results suggest that the severity of depth conflict problems might depend on the task characteristics and the UI design. It also shows that Dynamic Depth is a promising solution and was most preferred by our participants for video subtitles. Our results also show that Halo Blur did not work as expected in a dynamic VR video environment.

ACKNOWLEDGEMENT

We thank the user study participants for their time and feedback. Figure 1, 2, and 3 use images from YouTube users Kevin Kunze under a Creative Commons license. This work was supported in part by NSF IIS-1321119.

REFERENCES

1. 2017. Binocular Vision, Stereoscopic Imaging and Depth Cues. (2017). https://developer.oculus.com/design/latest/concepts/bp_app_imaging/
2. Robert Anderson, David Gallup, Jonathan T. Barron, Janne Kontkanen, Noah Snavely, Carlos Hernández, Sameer Agarwal, and Steven M. Seitz. 2016. Jump: virtual reality video. *ACM Transactions on Graphics* 35, 6 (nov 2016), 1–13. DOI: http://dx.doi.org/10.1145/2980179.2980257
3. Stephen R Arnott and Judith M Shedden. 2000. Attention switching in depth using random-dot autostereograms: Attention gradient asymmetries. *Attention, Perception, & Psychophysics* 62, 7 (2000), 1459–1473.
4. Blaine Bell, Steven Feiner, and Tobias Höllerer. 2001. View management for virtual and augmented reality. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*. ACM, 101–110.
5. Anastasia Bezerianos, Pierre Dragicevic, and Ravin Balakrishnan. 2006. Mnemonic Rendering: An Image-Based Approach for Exposing Hidden Changes in Dynamic Displays. In *Proceedings of the 19th annual ACM symposium on User interface software and technology - UIST '06*. ACM Press, New York, New York, USA, 159. DOI: http://dx.doi.org/10.1145/1166253.1166279
6. Tobias Blum, Matthias Wieczorek, Andre Aichert, Radhika Tibrewal, and Nassir Navab. 2010. The effect of out-of-focus blur on visual discomfort when using stereo displays. In *2010 IEEE International Symposium on Mixed and Augmented Reality*. IEEE, 13–17. DOI: http://dx.doi.org/10.1109/ISMAR.2010.5643544
7. David K. Broberg. 2011. Infrastructures for Home Delivery, Interfacing, Captioning, and Viewing of 3-D Content. *Proc. IEEE* 99, 4 (apr 2011), 684–693. DOI: http://dx.doi.org/10.1109/JPROC.2010.2092390
8. Jacqueline Chu, Chris Bryan, Min Shih, Leonardo Ferrer, and Kwan-Liu Ma. 2017. Navigable Videos for

Presenting Scientific Data on Affordable Head-Mounted Displays. In *Proceedings of the 8th ACM on Multimedia Systems Conference - MMSys'17*. ACM Press, New York, New York, USA, 250–260. DOI: http://dx.doi.org/10.1145/3083187.3084015

9. Bleeding Disorders Community. 2017. Factor Treatment in Hemophilia A and B. Video. (9 June 2017). Retrieved August 22, 2017 from https://www.youtube.com/watch?v=tHbFY5G3GVs.
10. Joao L C Dasilva. 2017. period film scene full VR 360 3D on film set. Video. (2 September 2017). Retrieved August 22, 2017 from https://www.youtube.com/watch?v=nH1xJox61z4.
11. IRALTA VR Productora de video. 2016. 360 3D The Circus School CARAMPA. Video. (19 February 2016). Retrieved August 22, 2017 from https://www.youtube.com/watch?v=tABJlVww-2M.
12. devinsupertramp. 2017. Terrifying Masquerade Party in 3D 360!! Video. (8 February 2017). Retrieved August 22, 2017 from https://www.youtube.com/watch?v=mb_J_bMor1g.
13. Piotr Didyk, Tobias Ritschel, Elmar Eisemann, Karol Myszkowski, Hans-Peter Seidel, and Wojciech Matusik. 2012. A luminance-contrast-aware disparity model and applications. *ACM Transactions on Graphics* 31, 6 (nov 2012), 1. DOI: http://dx.doi.org/10.1145/2366145.2366203
14. Adeola Fabola, Alan Miller, and Richard Fawcett. 2015. Exploring the past with Google Cardboard. In *Digital Heritage, 2015*, Vol. 1. IEEE, 277–284.
15. Steven Feiner, Blair MacIntyre, Marcus Haupt, and Eliot Solomon. 1993. Windows on the world: 2D windows for 3D augmented reality. In *Proceedings of the 6th annual ACM symposium on User interface software and technology*. ACM, 145–155.
16. Davide Gadia, Marco Granato, Dario Maggiorini, Laura Anna Ripamonti, and Cinzia Vismara. 2017. Consumer-oriented Head Mounted Displays: Analysis and Evaluation of Stereoscopic Characteristics and User Preferences. *Mobile Networks and Applications* February (feb 2017), 1–11. DOI: http://dx.doi.org/10.1007/s11036-017-0834-9
17. ZDF Enterprises GmbH. 2017. Gladiators In The Roman Colosseum VR 3D 360°. Video. (28 February 2017). Retrieved August 22, 2017 from https://www.youtube.com/watch?v=XBuijx_iZtQ.
18. David M Hoffman and Martin S Banks. 2010. Focus information is used to interpret binocular images. *Journal of vision* 10, 5 (2010), 13. DOI: http://dx.doi.org/10.1167/10.5.13
19. Hong Hua. 2017. Enabling Focus Cues in Head-Mounted Displays. *Proc. IEEE* 105, 5 (may 2017), 805–824. DOI: http://dx.doi.org/10.1109/JPROC.2017.2648796
20. Jingwei Huang, Zhili Chen, Duygu Ceylan, and Hailin Jin. 2017. 6-DOF VR videos with a single 360-camera. In *2017 IEEE Virtual Reality (VR)*. IEEE, 37–44. DOI: http://dx.doi.org/10.1109/VR.2017.7892229
21. Robert Kennedy, Norman Lane, Kevin Berbaum, and Michael Lilienthal. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. (1993).
22. Johannes Kopf. 2016. 360 video stabilization. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 195.
23. Gregory Kramida and Amitabh Varshney. 2015. Resolving the Vergence-Accommodation Conflict in Head Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics* 22, 7 (2015), 1–16. DOI:http://dx.doi.org/10.1109/TVCG.2015.2473855
24. Arun Kulshreshth and Joseph J. LaViola. 2016. Dynamic Stereoscopic 3D Parameter Adjustment for Enhanced Depth Discrimination. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. ACM Press, New York, New York, USA, 177–187. DOI: http://dx.doi.org/10.1145/2858036.2858078
25. Jeremy Lacoche, Morgan Le Chenechal, Sebastien Chalme, Jerome Royan, Thierry Duval, Valerie Gouranton, Eric Maisel, and Bruno Arnaldi. 2015. Dealing with frame cancellation for stereoscopic displays in 3D user interfaces. In *2015 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 73–80. DOI: http://dx.doi.org/10.1109/3DUI.2015.7131729
26. M. Lambooij, M.J. Murdoch, W.A. IJsselsteijn, and I. Heynderickx. 2013. The impact of video characteristics and subtitles on visual comfort of 3D TV. *Displays* 34, 1 (jan 2013), 8–16. DOI: http://dx.doi.org/10.1016/j.displa.2012.09.002
27. Marc T. M. Lambooij, Wijnand A. IJsselsteijn, and Ingrid Heynderickx. 2007. Visual discomfort in stereoscopic displays: a review. 6490, May 2010 (2007), 64900I. DOI: http://dx.doi.org/10.1117/12.705527
28. Robert S. Laramee and Colin Ware. 2002. Rivalry and interference with a head-mounted display. *ACM Transactions on Computer-Human Interaction* 9, 3 (sep 2002), 238–251. DOI: http://dx.doi.org/10.1145/568513.568516
29. Andrew MacQuarrie and Anthony Steed. 2017. Cinematic virtual reality: Evaluating the effect of display type on the viewing experience for panoramic video. In *Virtual Reality (VR), 2017 IEEE*. IEEE, 45–54.
30. Kevin Matzen, Michael F. Cohen, Bryce Evans, Johannes Kopf, and Richard Szeliski. 2017. Low-cost 360 stereo photography and video capture. *ACM Trans. Graphics* (2017).
31. Ming Hou. 2003. A model of real-virtual object interactions in stereoscopic augmented reality environments. In *Proceedings on Seventh International Conference on Information Visualization, 2003. IV 2003.*, Vol. 2003-Janua. IEEE Comput. Soc, 512–517. DOI: http://dx.doi.org/10.1109/IV.2003.1218033

32. Cuong Nguyen, Stephen DiVerdi, Aaron Hertzmann, and Feng Liu. 2017a. CollaVR: Collaborative In-Headset Review for VR Video. In *ACM symposium on User interface software and technology - UIST '17*. ACM.

33. Cuong Nguyen, Stephen DiVerdi, Aaron Hertzmann, and Feng Liu. 2017b. Vremiere: In-Headset Virtual Reality Video Editing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. ACM Press, New York, New York, USA, 5428–5438. DOI: http://dx.doi.org/10.1145/3025453.3025675

34. Thomas Oskam, Alexander Hornung, Huw Bowles, Kenny Mitchell, and Markus Gross. 2011. OSCAM - optimized stereoscopic camera control for interactive 3D. In *Proceedings of the 2011 SIGGRAPH Asia Conference on - SA '11*. ACM Press, New York, New York, USA, 1. DOI:http://dx.doi.org/10.1145/2024156.2024223

35. Javier Sánchez Pérez, Enric Meinhardt-Llopis, and Gabriele Facciolo. 2013. TV-L1 optical flow estimation. *Image Processing On Line* 2013 (2013), 137–150.

36. Stephan Reichelt, Ralf Häussler, Gerald Fütterer, and Norbert Leister. 2010. Depth cues in human visual perception and their realization in 3D displays. In *Proc. SPIE*, Vol. 7690. 76900B.

37. Leila Schemali and Elmar Eisemann. 2014. Design and evaluation of mouse cursors in a stereoscopic desktop environment. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 67–70. DOI: http://dx.doi.org/10.1109/3DUI.2014.6798844

38. Jonas Schild, Liane Bölicke, Joseph J. LaViola Jr., and Maic Masuch. 2013. Creating and analyzing stereoscopic 3D graphical user interfaces in digital games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*. ACM Press, New York, New York, USA, 169. DOI: http://dx.doi.org/10.1145/2470654.2470678

39. Jonas Schild, Joseph J. LaViola, and Maic Masuch. 2014. Altering gameplay behavior using stereoscopic 3D vision-based video game design. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*. ACM Press, New York, New York, USA, 207–216. DOI: http://dx.doi.org/10.1145/2556288.2557283

40. Aljoscha Smolic, Peter Kauff, Sebastian Knorr, Alexander Hornung, Matthias Kunter, M MuÌĹller, and Manuel Lang. 2011. Three-Dimensional Video Postproduction and Processing. *Proc. IEEE* 99, 4 (apr 2011), 607–625. DOI: http://dx.doi.org/10.1109/JPROC.2010.2098350

41. Filippo Speranza, Wa J Tam, Ron Renaud, and Namho Hur. 2006. Effect of disparity and motion on visual comfort of stereoscopic images. In *Proceedings of SPIE: Stereoscopic Displays and Virtual Reality Systems XIII*, Andrew J. Woods, Neil A. Dodgson, John O. Merritt, Mark T. Bolas, and Ian E. McDowall (Eds.), Vol. 6055. 60550B. DOI:http://dx.doi.org/10.1117/12.640865

42. Wa James Tam, Filippo Speranza, Carlos Vázquez, Ron Renaud, and Namho Hur. 2012. Visual comfort: stereoscopic objects moving in the horizontal and mid-sagittal planes. In *Proc. SPIE*, Andrew J. Woods, Nicolas S. Holliman, and Gregg E. Favalora (Eds.), Vol. 8288. 828813. DOI:http://dx.doi.org/10.1117/12.909121

43. Junle Wang, Patrick Le Callet, Sylvain Tourancheau, Vincent Ricordel, and Matthieu Perreira Da Silva. 2012. Study of depth bias of observers in free viewing of still stereoscopic synthetic stimuli. *Journal of Eye Movement Research* 5, 5 (2012).

44. Colin Ware. 1995. Dynamic stereo displays. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '95* (1995), 310–316. DOI: http://dx.doi.org/10.1145/223904.223944

45. Yong Jung, Hosik Sohn, Seong-il Lee, and Yong Ro. 2014. Visual comfort improvement in stereoscopic 3D displays using perceptually plausible assessment metric of visual comfort. *IEEE Transactions on Consumer Electronics* 60, 1 (feb 2014), 1–9. DOI: http://dx.doi.org/10.1109/TCE.2014.6780918