This is the accepted manuscript of the article, which has been published in Rosenberg E. et al. (eds) 2017 IEEE Virtual Reality (VR). IEEE. ISBN: 978-1-5090-6647-6. ISSN: 2375-5334. https://doi.org/10.1109/VR.2017.7892285

Interaction with WebVR 360° Video Player: Comparing Three Interaction Paradigms

Toni Pakkanen*[1] Jaakko Hakulinen[1] Tero Jokela[2] Ismo Rakkolainen[1]
Jari Kangas[1] Petri Piippo[2] Roope Raisamo[1] Marja Salmimaa[2]

[1] School of Information Sciences, University of Tampere, Finland
[2] Nokia Technologies, Nokia Corporation, Finland

Abstract

Immersive 360° video needs new ways of interaction. We compared three different interaction methods to find out which one of them is the most applicable for controlling 360° video playback. The compared methods were: remote control, pointing with head orientation, and hand gestures. A WebVR-based 360° video player was built for the experiment.

Keywords: Immersive 360° video, interaction methods, gestures.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems — Artificial, augmented, and virtual realities; H.5.2 [User Interfaces]: Input devices and strategies; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism — Virtual reality

1 Introduction

In 360° video playback with VR headsets, the user is immersed in the video. This creates challenges for designing interactions for controlling the playback, as the user cannot see the real environment or even their hands. We compared three widely used VR interaction methods for controlling 360° video playback: hand gestures, pointing with head orientation, and remote control. An online web-based player, created by us, was used to allow a browser to play back 360° video over the Internet. A similar approach has been suggested for streaming Omni-Directional Video (ODV) [8].

The users prefer to use gestures for the interaction [2], but gesture-based user interfaces can be unreliable [1]. Thus, other methods, such as head orientation pointing, have been common in VR environments. Remote controls and tablets have been more successful on interaction with smart TVs than gestures [3]. The most commonly needed functions are skipping content and pausing the playback [6]. Also panning and zooming are among the most wished features [10]. It can be assumed that the same results would apply also for 360° video playback in a VR environment.

When asked to invent hand gestures, users tend to suggest single-hand gestures where the hand form does not change [5]. This supports an approach of using one-handed gestures, where the hand shape is stationary. The gestures that we chose for the experiment are among the most commonly used gestures in current applications [4]. This made our selection of gestures more familiar and easier to learn for the users. When users defined gestures for a dome-based ODV system, they suggested grabbing and dragging the image for panning and moving hand to left and right for rewind/fast forward [9]. As we used a Leap Motion sensor, which has a smaller recognition area, we used similar movement based clockwise and counterclockwise rotation gestures for rewind/fast forward instead. Since pointing is perceived as a difficult action to do with gestures [7], we chose static pose gestures instead.

In our approach, we used one-handed gestures, which had stationary hand postures, with small hand movements to initiate actions. The chosen gestures are commonly used and similar to the gestures users have been reported to suggest for similar tasks. The remote control and head pointing user interfaces included in the study are currently the industry de-facto standard interaction methods for 360° video playback. In this paper, we compare these three interaction methods with performance and usability measures.

* {firstname.lastname} [1]@sis.uta.fi, [2]@nokia.com

2 System

We built an immersive 360° video player software in JavaScript utilizing the WebGL and WebVR APIs via the ThreeJS library[1]. During the experiments, the software was run on a nightly build version of the Mozilla Firefox browser. The hardware setup consisted of an Oculus Rift DK2 VR headset, a Logitech media remote control, and a Leap Motion sensor for gesture control.

Figure 1: UI elements for gestures (left), head pointing (middle), and remote control (right).

In addition to controlling the direction of the video view with head orientation, the functionality of the player software consisted of play, pause, rewind, and fast forward commands for the video playback and panning the video view horizontally. The gesture control modes were: tapping with index finger to pause and play the video, pinching with index finger and thumb to grab and pan the view, and circular hand movement with index and middle finger extended to seek video (5 seconds per rotation). In remote control, relevant remote control buttons were used, with rewind and fast forward buttons having a dual function: a single click started a rewind/fast forward operation or increased the speed up to 16x rate, while holding the button for at least one second rewinded/fast forwarded until the button was released. In the head pointing method, a single remote control button was used to select actions.

[1] http://threejs.org/

A visual user interface overlay (Figure 1) appeared on top of the video when the user interacted with the system, that is, when a hand was detected by the Leap Motion sensor or after a button was pressed in the remote control and head pointing methods. The user interface appeared in the direction where the user was looking at when it was activated. In the gestures method, a hand cursor matching the recognized gesture was displayed whenever a user's hand was detected in the interaction area. In the head pointing method, a circular cursor was shown at the center of the view. If the user was pointing at a button, the button was enlarged to indicate the target. When the button was selected, its colors were highlighted to indicate the selection.

3 EXPERIMENT

18 voluntary participants experimented with the system. They executed eight tasks with each interaction method:

- Four different tasks for rewinding/fast forwading: twice to a certain time point, once by a certain amount of time, and once to a given visual point in the clip.
- One task for pause and one for continuing the playback
- Two tasks for panning the video image horizontally: one with paused video and another while the video was playing

Video clips consisted of scenes recorded around the city. The order of the interaction methods was counter-balanced between the participants and the order of the video clips was also randomized. The following data was collected for each interaction method:

- Task accuracy and completion times
- NASA-TLX Questionnaire
- Simulation Sickness Questionnaire (SSQ)
- Subjective evaluation with nine-point bipolar scales (-4 to 4) to rate the perceived accuracy, difficulty, pleasantness, distraction to watching the video, expectations, and efficiency.

4 RESULTS

Statistically significant results included the following:

Task times: 1) Rewind/fast forward was faster with both pointing and remote control than with gestures. 2) Panning the view was faster with remote control than with gestures.

SSQ: The participants had less nausea after the second time of using the VR headset than before it. Thus, after getting used to the VR headset, the participants felt better.

NASA-TLX: Gestures were perceived more demanding than 1) pointing in mental, physical, effort, and frustration, and 2) remote control in mental, physical, temporal, performance, effort, and frustration. Pointing was perceived worse than remote control in temporal demand.

Usability measures: Gestures were rated lower than 1) pointing and remote control in perceived accuracy with rewind/fast forward, difficulty, expectations and efficiency, and 2) remote control in accuracy in panning the view and pleasantness. Pointing was perceived less accurate in rewind/fast forward task than remote control.

Preference: Participants preferred remote control over gestures.

5 DISCUSSION AND FUTURE WORK

Based on the results in performance and traditional usability attributes that we experimented with, the best interaction method for 360° video in a VR display environment would be either a well-shaped remote control or pointing using head orientation with a graphical VR user interface.

At the current level of technology, gestures are problematic and more mature technology and research are necessary to find efficient gestures to use in 360° video environments with VR headsets. Our results show that gestures were slower to use than the other two interaction methods. Also based on the NASA-TLX results, gestures were considered more demanding than the other two interaction methods. In terms of usability, gestures were perceived worse than pointing and remote control. Between pointing and remote control there were small differences for remote control's benefit. Thus, a familiar and well-known remote control was found to be the best, even when it cannot be seen.

The participants' dissatisfaction with the gestures method might be partly explained by problems in gesture recognition. First, participants had difficulties in keeping their hand within the recognition area of the Leap Motion sensor. This observation suggests a need for improvements in sensor design, as the gesture recognition systems for VR environments would need a larger input area. The second observation during the experiments was that when the participants encountered a problem with gestures, they had a tendency to try to do larger and faster gestures, which resulted in more problems. The participants had most problems with the rewind/fast forward gestures. These observations call for efficient, simple, and easily recognizable gestures, which the users can perform blindly. Finding the optimal combination of different interaction methods for different tasks would also be needed.

REFERENCES

[1] R. Atienza, R. Blonna, M. I. Saludares, J. Casimiro, and V. Fuentes, "Interaction techniques using head gaze for virtual reality," *2016 IEEE Region 10 Symposium (TENSYMP),* Bali, 2016, pp. 110-114.

[2] L. Bleumers, W. van den Broeck, B. Lievens, and J. Pierson. 2012. Seeing the bigger picture: a user perspective on 360° TV. In *Proceedings of the 10th European conference on Interactive tv and video (EuroiTV '12).* ACM, New York, NY, USA, 115-124.

[3] J. Bobeth, J. Schrammel, S. Deutsch, M. Klein, M. Drobics, C. Hochleitner, and M. Tscheligi. 2014. Tablet, gestures, remote control?: influence of age on performance and user experience with iTV applications. In *Proceedings of the 2014 ACM international conference on Interactive experiences for TV and online video (TVX '14).* ACM, New York, NY, USA, 139-146.

[4] A. Cabreira and F. Hwang. 2015. An analysis of mid-air gestures used across three platforms. In *Proceedings of the 2015 British HCI Conference (British HCI '15).* ACM, New York, NY, USA, 257-258.

[5] D. Grijincu, M. Nacenta, and P. Kristensson. 2014. User-defined Interface Gestures: Dataset and Analysis. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces (ITS '14).* ACM, New York, NY, USA, 25-34.

[6] M. Darnell. How do people really interact with TV?: naturalistic observations of digital tv and digital video recorder users. *Comput. Entertain. 5*, 2, Article 10. 2007.

[7] P. Ponsa, C. Urbina, C. Manresa-Yee, and R. Vilanova. A Usability Study Case of a Vision-Based Gesture Interface. In *Proceedings of the XV International Conference on Human Computer Interaction (Interacción '14).* ACM, New York, NY, USA, Article 26. 2014.

[8] P. Quax, J. Liesenborgs, P. Issaris, W. Lamotte, and J. Claes. 2013. A practical and scalable method for streaming omni-directional video to web users. In *Proceedings of the 2013 ACM international workshop on Immersive media experiences (ImmersiveMe '13).* ACM, New York, NY, USA, 57-60.

[9] G. Rovelo Ruiz, D. Vanacken, K. Luyten, F. Abad, and E. Camahort. 2014. Multi-viewer gesture-based interaction for omni-directional video. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14).* ACM, New York, NY, USA, 4077-4086.

[10] G. Zoric, L. Barkhuus, A. Engström, and E. Önnevall. 2013. Panoramic video: design challenges and implications for content interaction. In *Proceedings of the 11th european conference on Interactive TV and video (EuroITV '13).* ACM, New York, NY, USA, 153-162.