A SUBJECTIVE STUDY OF VIEWER NAVIGATION BEHAVIORS WHEN WATCHING 360-DEGREE VIDEOS ON COMPUTERS

Fanyi Duanmu[1], Yixiang Mao[1], Shuai Liu[1], Sumanth Srinivasan[2], and Yao Wang[1]
[1]New York University, Brooklyn, NY 11201, USA; [2]Vimeo Inc., New York, NY 10011, USA.
[1]{fanyi.duanmu, yixiang.mao, shuai.liu, yaowang}@nyu.edu; [2]sumanth@vimeo.com

ABSTRACT

Virtual reality (VR) applications have become popular recently and rapidly commercialized. The behaviors of users watching 360-degree omni-directional videos have not been fully investigated. In this paper, a dataset of view trajectories for users watching 360-degree videos over computer (desktop/laptop) environment is presented. The dataset includes view center trajectory data collected from viewers watching twelve 360-degree videos over the computer, using mouse to navigate and explore the environment. The selected videos cover a variety of contents, leading to different navigation patterns and behaviors. Based on the dataset, we demonstrate that the viewers share similar viewing patterns over certain 360-degree video category. We also compare the view motion patterns and statistics with prior datasets captured using head-mounted display (HMD). The dataset has been made available online, to facilitate the studies on 360-degree video view prediction, content saliency analysis and VR streaming, etc.

Index Terms— 360-degree Video, Video Dataset, Virtual Reality, Head Motion, Focus of Attention.

1. INTRODUCTION

1.1. Motivation

360-degree video has become popular in recent years due to the technological advances in virtual reality (VR) and augmented reality (AR) technologies and has been rapidly commercialized in many applications, such as immersive cinema, gaming, education, healthcare, social media and 360-degree video streaming, etc. Beyond traditional 2D videos, 360-degree video provides more engaging and immersive experience. Currently, a number of companies are working towards developing new VR/AR services and products. For example, HTC Vive [1] and Oculus Rift [2] are the leading high-end VR display system manufacturers on the market. The VR streaming services are already made available on several major platforms such as YouTube [3], Facebook [4], etc.

Unlike traditional 2D video, in 360-degree video, only a portion of the entire scene is watched at a time and users keep exploring and navigating the view direction according to the contents and viewers' interests. However, the viewing behaviors of 360-degree video have not been investigated thoroughly. Many open questions remain, such as how people react to the virtual environment and target at the regions of interest (ROI), how the viewing patterns change from time to time according to the VR content variation, what are the typical head motions while watching a particular category of 360-degree video. These questions are far more difficult to understand than traditional 2D video due to the introduction of the new dimension of view direction and therefore require more sophisticated content-level analyses and view pattern studies. Understanding how users react and navigate in 360-degree environments will benefit many applications, such as 360-degree video coding and streaming (in which the ROI regions can be coded with higher bit-allocation and streamed with higher priority while non-ROI regions can be processed coarsely or even not included to reduce the bandwidth, using tile-based or tier-based solutions [5]-[9]), VR content design (gaming, media, cinema, etc.), psychological studies, etc. To facilitate both the industrial and academic communities towards increasing adoptions of VR and AR technologies in the future, it is therefore desirable to have 360-degree video datasets for different types of contents captured with different display configurations. In this paper, a novel dataset collected from viewers watching 360-degree videos over computer environment is proposed.

1.2. Previous Work

There are a few 360-degree video view trace datasets proposed previously. In [10], a 360-degree image dataset is proposed for user navigation pattern exploration and saliency map generation. The dataset contains trajectory information of 32 users over 21 test images and is made public. In [11], a testbed is developed to conduct 360-degree video quality assessments via HMD. Average viewing probabilities are derived during the subjective tests. This work is targeted at image level subjective evaluation and the test samples are relatively limited, i.e.,

978-1-5386-1737-3/18/$31.00 ©2018 IEEE

6 images. In [12], a 360-degree video coding and quality evaluation framework is proposed along with a 10-video dataset released. Head motion trajectories are combined to derive a weighted PSNR and the viewing probability map is reported. However, the subject number is limited, i.e., 10 participants, and the navigation patterns for individual participant and corresponding video contents are not provided publicly. There are several 360-degree video datasets published recently. For example, in [13], a 360-degree video head movement dataset containing 59 users watching five 70-second videos is provided. The data is provided in 3-degree freedom without considering the user translational movements. Head motion statistics (e.g., angular distance, angular velocity, attendance CDF, etc.) are provided and analyzed. In [14], a similar HMD-based head motion dataset is proposed, consisting of 48 users watching 18 spherical videos in 5 categories. The data is documented in 6-degree freedom. Statistical analysis and data visualization are provided. In [15], another HMD dataset is proposed with viewer orientation data captured from 50 subjects watching 10 videos. Both the viewer orientation data and the detected saliency maps are provided publicly.

1.3. Our Contributions

Firstly, we create a unique dataset containing viewers' trajectories when watching the 360-degree videos over desktop/laptop environment. To our best knowledge, this is the first available computer-based 360-degree video view trajectory dataset.

Secondly, we conduct statistical studies and pattern analysis over the traces, to explore the viewer behaviors when watching 360-degree video over computer and navigating the viewport using mouse and investigate how these viewing patterns are correlated to the contents.

Thirdly, we compare and analyze the viewing patterns and statistics with those reported in the prior works, in which viewers watch 360-degree videos through HMDs, to explore the similarities and differences between HMD-based and computer-based VR navigations.

The rest of the paper is structured as follows. Section 2 introduces our dataset, including the data collection procedures, the dataset structure and format interpretation. Section 3 provides the view motion analysis and the trace data visualization. Section 4 concludes this paper with future work summarized.

2. DATASET

2.1. 360-degree Video Selection

We selected twelve 360-degree videos downloaded from YouTube as our testing materials, as summarized in Table I. For each video, we select a representative clip of approximately 1-2 minutes. These selected sample videos cover a variety of video contents, including virtual tour, VR gaming, stage-performance, sports, movies, etc., to trigger content-dependent user navigation behaviors. The source videos are downloaded in equi-rectangular format in the highest available resolution and bitrate provided by YouTube. Sample frames are provided in Figure 1.

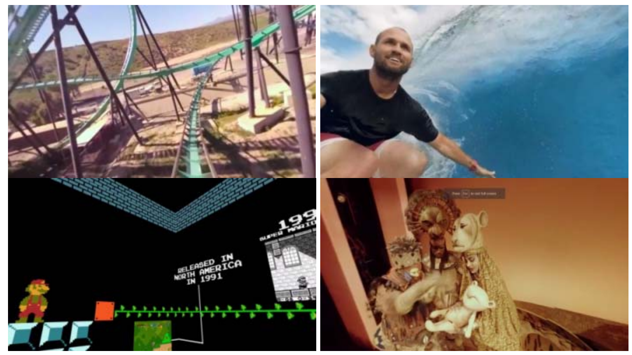

Figure 1. Sample 360-degree video frames after viewport rendering

2.2 Data Collection Procedures

We developed a 360-degree video streaming platform based on [16] to collect viewers' navigation trajectories. The video contents are archived under a local server. For each participant, a video streaming session is created via a local intranet link to ensure the video display continuity. When the participants navigate the viewport direction, the mouse translations and the corresponding timestamps are documented at the viewer's local cache, as detailed in Section 2.4. Later, when the video session is completed (i.e., when the user presses "Next" or "Finish" button), the data log will be up-streamed to our server and stored.

At the beginning of the test, an instructional page is provided to describe the experiments and procedures (e.g., how to navigate the viewport, how to pause and proceed, etc.). The participants are reminded to view in full-screen and wear headset for better quality of experience (QoE) and a more accurate trace generation. A sample video-clip is provided as a "training" video for the participants to get familiar with the environment and the mouse sensitivity. The viewers can freely pause in between two adjacent videos and proceed when ready.

Totally we have over 50 viewers participated in the data collection. The participants are university scholars and students of age ranging from 20 to 35. From the collected traces, we further visualize the data and remove the "outliers" with visibly irregular navigation patterns (e.g., video rewinding) or have irrationally long periods of inactivity. In our subjective tests, we do not restrict viewers to follow specific viewing distance setup or lighting conditions, with the hope that the captured viewing statistics are general reflections under free-viewing conditions. Our main intention is to explore how the viewer navigation behaviors are driven by the 360-degree video contents.

Table I. Test 360-degree Video Metadata and Description

Video Name	YouTube Suffix	Category	Resolution	Start Offset	Duration	Content Description
Surfing	7gjR60TSn8Q	Sport	3840x2048	1:00	1:15	A group of travelers diving and surfing in the water.
Wing Suit	t99N223fqCo	Sport	3840x2048	0:00	1:13	Extreme sport with people wearing wingsuit and gliding in the sky.
Skiing	0wC3x_bnnps	Sport	3840x2048	0:40	1:00	Sport shooting of a man skiing with a mounted 360° camera.
Mega Coaster	xNN-bJQ4vI	Sport	3840x2048	0:30	1:20	Roller coaster experience with a mounted 360° camera.
Paris	sJxiPiAaB4k&t	Tour	3840x2048	0:00	1:00	Virtual tour of Paris with fixed camera.
Time Lapse	CIw8R8thnm8	Tour	3840x2048	0:00	1:31	New York Sightseeing of famous landscapes with fixed camera.
Amsterdam	FzrkpXlRP1M	Tour	3840x2048	1:32	1:17	Virtual tour of Amsterdam with fixed camera.
Spotlight Help	G-XZhKqQAHU	Movie	3840x2048	1:00	1:30	A movie story of a monster chasing civilian.
Lion King	7T57kzGQGto	Show	3840x2048	3:30	1:22	Virtual stage-show of Lion King opera with scene transitions.
Clan	wczdECcwRw0	Game	2048x1024	0:00	1.23	Game video of elf defending their castle with fixed camera.
Super Mario	OWl_dkjHiF8	Game	3840x2048	1:12	1:42	Game video of Super Mario.
Hog Rider	yVLfEHXQk08	Game	3840x2048	0:00	1:13	Game video of invaders riding hogs with a mounted 360° camera.

2.3 Dataset Structure

The captured view trajectory data is structured as illustrated in Figure 2. Under the root folder, a meta data log is provided with detailed video information, such as video resolution, frame rate, starting timestamp and video duration, etc. Underneath, separate folders are created to archive each individual video and are named according to Table I. In each video folder, user view motion traces captured from all participants are documented in ".csv" files with three columns for time, yaw and pitch values, respectively, as detailed in Section 2.4.

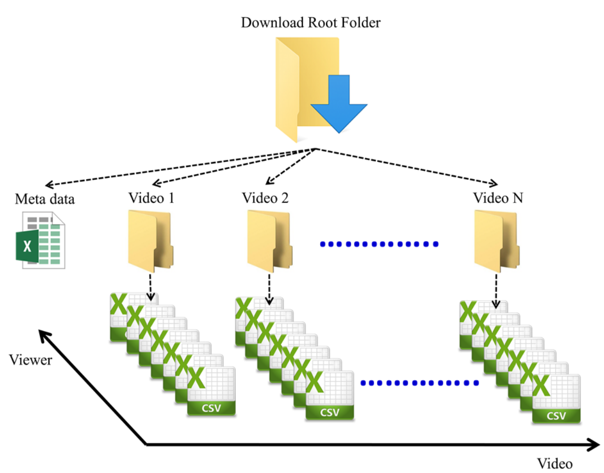

Figure 2. Dataset folder structure

2.4 Data Format

Our data is structured as a "time-yaw-pitch" triplet.

- Time: sampling timestamp relative to the display start time (i.e., 0 second).
- Yaw: horizontal translated viewport center in radians after circular rotation of 2π. The reported yaw values are between (-π, π] with 0 indicating the horizontal center of the original equi-rectangular video.
- Pitch: vertical translated viewport center in radians. The vertical movement is clamped within (-π/2, π/2] with 0 indicating the vertical center of the original equi-rectangular video.

Our dataset is made public and can be downloaded from http://videolab.engineering.nyu.edu/Desktop360.

3. DATA ANALYSIS AND VISUALIZATION

In this section, we conduct data analyses to understand how viewers explore different 360-degree video contents. We also provide simulation and visualization results to demonstrate the similarities and differences compared with head motion data captured via HMD.

3.1 View Center Distribution

Firstly, we track the view center locations of all the participants and analyze their distributions over different videos at different times. Figure 3 provides some sample results at two different timestamps and the sequence-level viewer focus distributions in heat-maps.

The experimental results coincide with the observations reported from the previous works, e.g., [13] [14]. Usually, at the beginning of the streaming session, viewers tend to explore the environment more vigorously and the viewer focuses are more scattered. Afterwards, the viewer focuses are more driven by the contents (e.g., story line, salient objects). For example, over "Mega Coaster" video, the viewers mostly concentrate on the central track and nearby areas (Figure 3, 1st row). In "Amsterdam" Scene 1 (Figure 3, 2nd row left), the viewers' attentions are distributed widely over buildings, vehicles and nearby pedestrians, since there are no dominant salient objects in the surrounding. However, in "Amsterdam" Scene 2, most viewers tend to focus on either the front view area where the boat is heading or the female guide (potentially driven by her audio instructions). Similarly, over the sport videos, such as "Skiing" (Figure 3, 3rd row), viewers are mainly driven by the front view and nearby skiers. Over gaming videos, such as "Super Mario" (Figure 3, 4th row), viewers basically eye-track the game character and follow the storyline. Our experimental results also validate the conclusion drawn from [10] that viewers are mostly focusing on the central regions rather than the polar areas (in the pitch direction of the equi-rectangular video format), as shown in the right column of Figure 3.

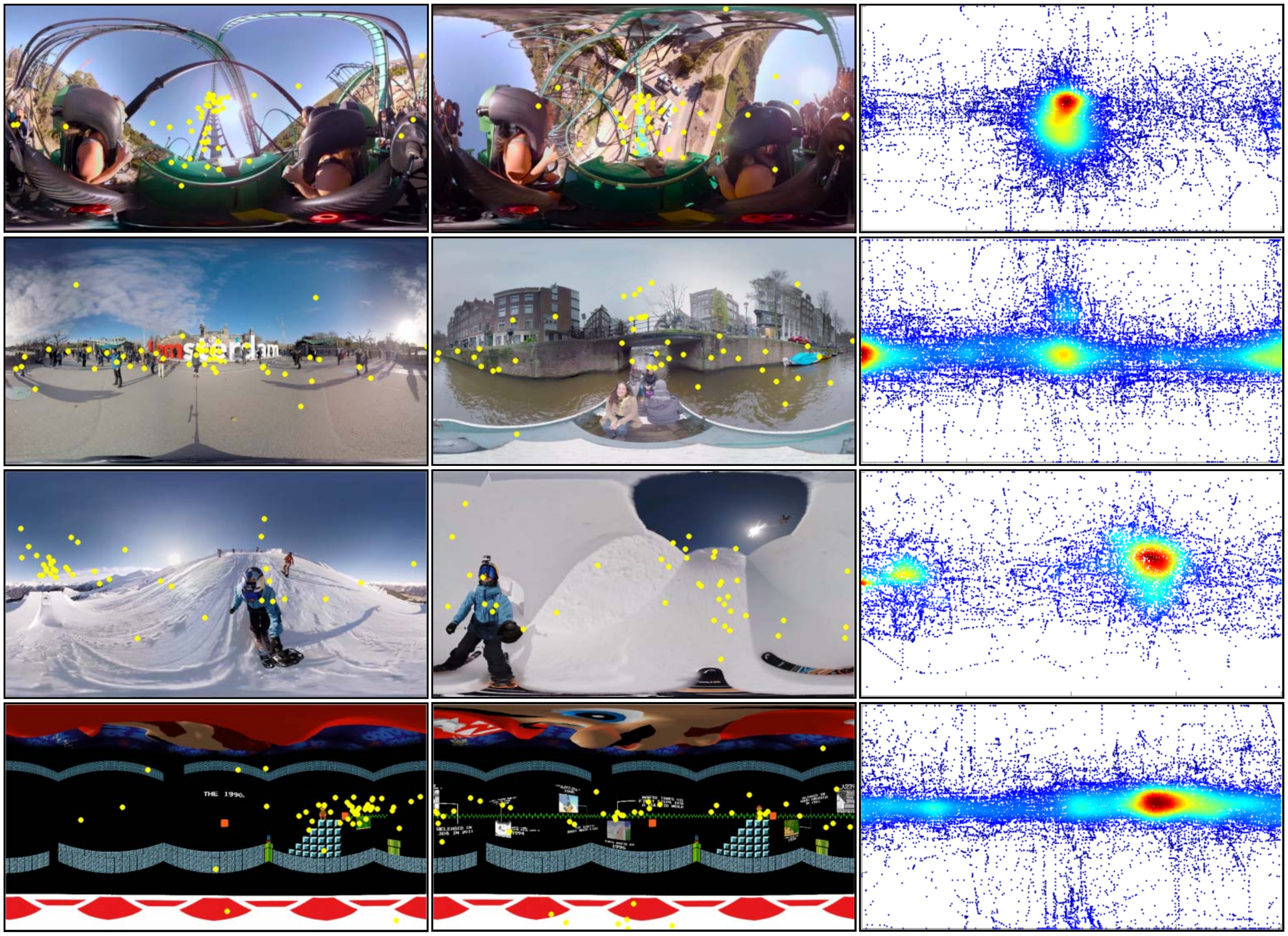

Figure 3. Focus Distribution Analyses over Different 360-degree Video Contents
First Row: "MegaCoaster"; Second Row: "Amsterdam"; Third Row: "Skiing"; Fourth Row: "SuperMario".
First and Second Columns: View center plots at different times; Third Column: View center plot over the entire video clip.

3.2 Trajectory Analysis

In Figure 4, sample viewer view center trajectories are provided, in both yaw and pitch directions. Please note that the data over the starting phase is excluded, in which users are randomly exploring the environment.

For 360-degree videos with salient objects or storyline, e.g., "MegaCoaster", "SuperMario", viewers' trajectories are bounded within a smaller dynamic range, whereas for 360-degree videos without such saliency clues, such as virtual tour sequences (e.g., "Paris", "Amsterdam"), the trajectories are more dynamic.

Besides, Figure 4 validates that the horizontal motions are more dominant than the vertical motions for most 360-degree videos. Viewers rarely look into the top or bottom views.

3.3 View Motion Statistics

To explore the motion differences between HMD and computer environments, we further calculate the angular movements over different segment lengths for different 360-degree videos. Please note that the extreme angular movement is π, i.e., the maximum angular distance the viewer can travel over the sphere.

In Figure 5, the cumulative density function (CDF) of the maximum angular movement is provided to show the distribution of the angular distances traveled within a 2-second window. From this distribution, we observe that 360-degree videos can be roughly classified into "slow view change" (SVC) and "fast view change" (FVC) categories. Empirically, we select a CDF threshold of 0.6 to cluster different 360-degree videos. For example, for a given video, at CDF = 0.6, if the translated angular distance is greater than $\pi/4$, the 360-degree video is classified as a FVC sequence, otherwise, a SVC sequence. As shown in Figure 5, for a 2-second segment, over 80% of viewers do not move the mouse more than $\pi/2$ radians from the initial position. This percentage is slightly lower than the statistics reported in [13] (i.e., 95%).

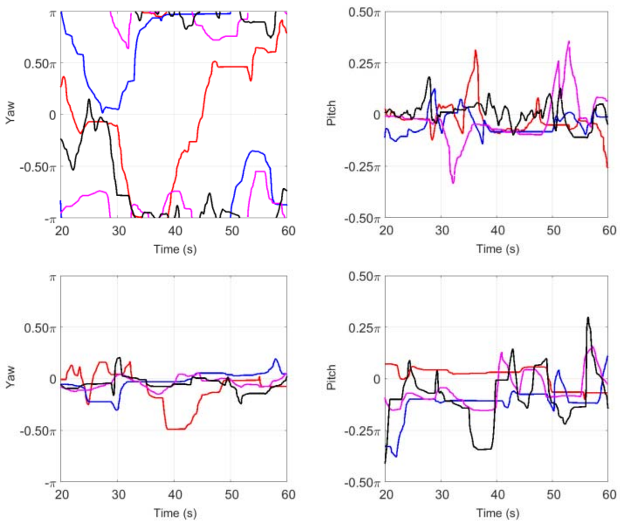

Figure 4. View Center Trajectory Plot
Top: "Amsterdam"; Bottom: "Mega Coaster". Each color corresponds to a specific participant.

Another interesting behavior is that the probability for "no-motion" zone (i.e., nearly zero angular distance) is quite large (over 20%), which is significantly higher than the statistics of HMD data. One explanation is that the 360-degree video usually starts with ROIs located in the view center and users tend to stay in this position longer. Another explanation is that HMD users have intrinsic small head movements, whereas over computer, the user can stay at a particular ROI stably for a longer period of time, when navigating using mouse.

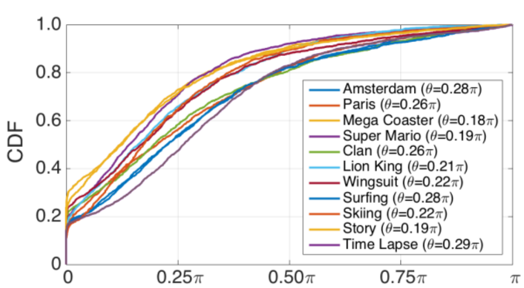

Figure 5. CDF of the maximum angular distance over 2-second for different 360-degree video contents. Average angular distance per video is reported in the legend.

Over the FVC and SVC 360-degree videos, we deduce the view motion statistics using different time scale of 1, 2, 3 and 5 seconds, respectively, as plotted in Figure 6. Over "Mega Coaster" (i.e., SVC video), 80% of viewers do not move beyond $\pi/2$ in a 5-second duration, whereas over "Time Lapse" (i.e., FVC video), >50% of viewers move beyond $\pi/2$ within the same duration.

In Figure 7, view motion angular velocity statistics over FVC video "Time Lapse" and SVC video "Mega Coaster" are presented. Each color represents a particular participant. The following conclusions can be drawn:

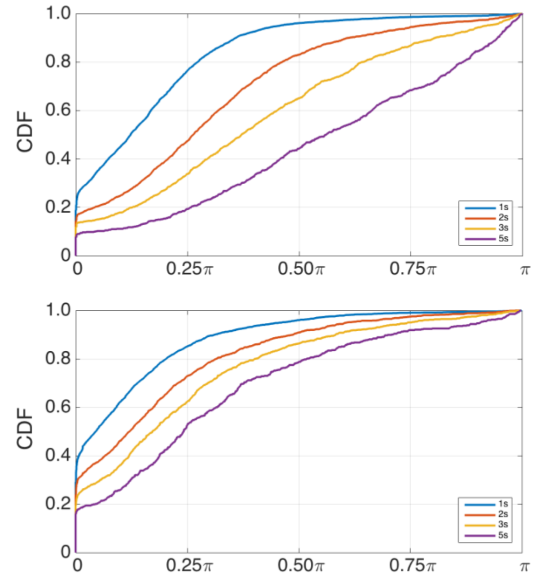

Figure 6. CDF of the maximum angular distance over different segment lengths (Top: "Time Lapse"; Bottom: "Mega Coaster")

Firstly, beyond the content stimulus, different viewer exhibits consistent navigation pattern. For example, the viewer who navigates faster or slower will mostly do the same across videos (e.g., Viewer 2 marked in green and Viewer 5 marked in yellow).

Secondly, viewers' motion patterns are driven by the contents. For example, for "Time Lapse", people tend to look around and continuously explore the environment. Therefore, the high-speed component is more dominant than "Mega Coaster", in which viewers mostly focus on the central track areas (as visualized in Figure 3).

Thirdly, compared with HMD, the motions triggered by mouse navigation contain a static zone, during which viewers do not move. This happens between the moment when a viewer releases the mouse and the next moment when the viewer re-clicks on the mouse to continue. In contrast, HMD viewers can seamlessly navigate the viewing direction, as discussed and visualized in [14]. Such mouse navigation leads to the large initial values in the CDF envelope in Figure 7. For example, in "Mega Coaster", Viewer 5 (marked in yellow) spends up to 60% time either not moving or only moving slightly (i.e., < 1 degree). For FVC videos (such as "Time Lapse"), most viewers keep exploring and therefore the CDF has lower initial values.

We further calculate the view velocity statistics, as summarized in Table II. Generally, "Sport" and "Tour" videos have larger mean velocity and viewers tend to continuously change the viewing direction and adjust the velocity (i.e., leading to a larger standard deviation). In contrast, "Movie", "Show" and "Game" videos are more context-driven, in which viewers tend to follow the storyline and therefore change the viewing direction more slowly and smoothly, therefore resulting in smaller mean velocity and standard deviation.

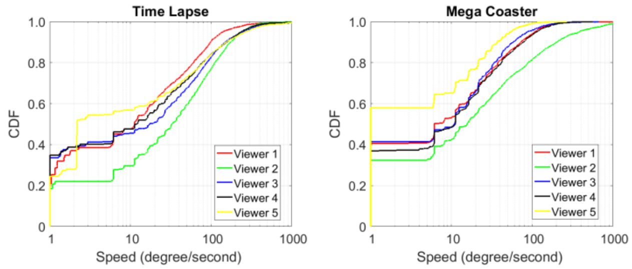

Figure 7. CDF of the Angular Velocity of View Center Motion

We also calculate the velocity statistics from the HMD datasets (i.e., [13] [14]) over the shared sequences. Experimental results demonstrate that HMD users tend to change their view centers faster and have more velocity variations than the viewers watching 360-degree videos over computer environment.

Table II. View Velocity Analysis

Video Name	Category	Velocity Average (deg/sec)	Velocity Standard Deviation (deg/sec)
Surfing	Sport	40.22	98.26
Wing Suit	Sport	36.43	119.03
Skiing*	Sport	33.25 (HMD) 26.52 (Computer)	146.39(HMD) 94.18(Computer)
Mega Coaster	Sport	28.83	89.03
Paris	Tour	32.40	67.18
Time Lapse*	Tour	41.60 (HMD) 36.54 (Computer)	272.87 (HMD) 117.90 (Computer)
Amsterdam	Tour	37.49	106.37
Spotlight Help*	Movie	28.21 (HMD) 19.79 (Computer)	212.94 (HMD) 82.60 (Computer)
Lion King	Show	26.18	55.04
Clan	Game	30.54	65.72
Super Mario	Game	24.60	56.24
Hog Rider	Game	23.13	68.23

Please note: videos marked with "*" are used in both prior HMD dataset and our dataset.

4. CONCLUSION

In this paper, a novel dataset is presented, containing the viewing center trajectories of users watching 360-degree videos over computer environment. The dataset is publicly released along with the source video clips. We present the data collection methodology, data format and also analyze the view motion statistics and compare with those reported in the prior HMD datasets.

We expect this dataset could benefit the researchers in areas of 360-degree video streaming, VR content creation and VR attention modeling. The future study includes the view prediction algorithm development for 360-degree video or VR streaming applications in the desktop/laptop environment.

REFERENCES

[1] HTC Vive. https://www.vive.com.
[2] Oculus Rift. https://www.oculus.com/rift.
[3] YouTube. http://www.youtube.com.
[4] Facebook. http://www.facebook.com/Facebook360.
[5] F. Duanmu, E. Kurdoglu, Y. Liu and Y. Wang, "View Direction and Bandwidth Adaptive 360 Degree Video Streaming using a Two-Tier System", in Proc. of IEEE International Symposium on Circuits and Systems, pp. 1-4, Baltimore, Maryland, USA, 2017.
[6] F. Duanmu, E. Kurdoglu, A. Hosseini, Y. Liu and Y. Wang, "Prioritized Buffer Control in Two-Tier 360 Video Streaming," in Proc. of ACM SigComm Workshop (ACM SigComm'17 VR/AR Network), pp. 13-18, Los Angeles, California, USA, 2017.
[7] L. Sun, F. Duanmu, Y. Liu, Y. Wang, Y. Ye, H. Shi, and D. Dai, "Multi-path Multi-tier 360-degree Video Streaming in 5G Networks," in ACM Multimedia System Conference (MMSys), Amsterdam, Netherland, 2018.
[8] Y.-K. Wang, Hendry, and M. Karczewicz, "Tile based VR video encoding and decoding schemes," JCTVC Doc. X0077, 2016.
[9] A. Zare, A. Aminlou, M. M Hannuksela, and M. Gabbouj, "HEVC-compliant tile-based streaming of panoramic video for virtual reality applications," In Proc. of ACM on Multimedia Conference. pp. 601–605, 2016.
[10] A. De Abreu, C. Ozcinar, and A. Smolic. "Look around you: Saliency maps for omnidirectional images in VR applications". In Proceedings of IEEE QoMEX, 2017.
[11] E. Upenik, M. Rerabek, and T. Ebrahimi. "A Testbed for Subjective Evaluation of Omnidirectional Visual Content". In 32nd Picture Coding Symposium, number EPFL-CONF-221560, 2016.
[12] M. Yu, H. Lakshman, and B. Girod. "A framework to evaluate omnidirectional video coding schemes". In Proc. of IEEE Mixed and Augmented Reality (ISMAR), 2015.
[13] Xavier Corbillon, Francesca De Simone, and Gwendal Simon, "360-Degree Video Head Movement Dataset," in Proc. of the 8th ACM on Multimedia Systems Conference (MMSys'17). ACM, New York, NY, USA, 199-204, 2017.
[14] Chenglei Wu, Zhihao Tan, Zhi Wang, and Shiqiang Yang, "A Dataset for Exploring User Behaviors in VR Spherical Video Streaming," In Proc. of the 8th ACM on Multimedia Systems Conference (MMSys'17). ACM, New York, NY, USA, 193-198, 2017.
[15] Ching-Ling Fan, Jean Lee, Wen-Chih Lo, Chun-Ying Huang, Kuan-Ta Chen, and Cheng-Hsin Hsu, "Fixation Prediction for 360° Video Streaming in Head-Mounted Virtual Reality," In Proc. of the 27th Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV'17). ACM, New York, NY, USA, 67-72, 2017.
[16] Simple-360-Player. https://github.com/slawrence/simple-360-player/tree/master/src.