A Dataset for Exploring User Behaviors in VR Spherical Video Streaming

Chenglei Wu, Zhihao Tan, Zhi Wang, Shiqiang Yang
Department of Computer Science and Technology, Graduate School at Shenzhen, Tsinghua University
{wcl15@mails.,tzh16@mails.,wangzhi@sz.,yangshq@}tsinghua.edu.cn

ABSTRACT

With Virtual Reality (VR) devices and content getting increasingly popular, understanding user behaviors in virtual environment is important for not only VR product design but also user experience improvement. In VR applications, the head movement is one of the most important user behaviors, which can reflect a user's visual attention, preference, and even unique motion pattern. However, to the best of our knowledge, no dataset containing this information is publicly available. In this paper, we present a head tracking dataset composed of 48 users (24 males and 24 females) watching 18 sphere videos from 5 categories. We carefully record how users watch the videos, how their heads move in each session, what directions they focus, and what content they can remember after each session. Based on this dataset, we show that people share certain common patterns in VR spherical video streaming, which are different from conventional video streaming. We believe the dataset can serve good resource for exploring user behavior patterns in VR applications.

CCS CONCEPTS

• **Information systems** → Multimedia databases; • **General and reference** → *Evaluation*;

KEYWORDS

Spherical Video; Head movement; Virtual reality

ACM Reference format:
Chenglei Wu, Zhihao Tan, Zhi Wang, Shiqiang Yang. 2017. A Dataset for Exploring User Behaviors in VR Spherical Video Streaming. In *Proceedings of MMSys'17, Taipei, Taiwan, June 20-23, 2017,* 6 pages.
DOI: http://dx.doi.org/10.1145/3083187.3083210

1 INTRODUCTION

With the launches of high-end virtual reality (VR) systems like HTC Vive and Oculus Rift, consumer-ready VR devices have made significant strides in the last year. The increasing popularity of VR devices has promoted the developments of various VR applications, including not only the traditional entertainment or educational applications (e.g., VR game), but also promising new applications such as live VR streaming, which can provide far more engaging experience than traditional live streaming. Researchers and engineers have also devoted their efforts to improve the user experience, e.g., improving the feeling of presence [6, 9] or reducing the motion sickness or discomfort caused by delay and 3D stereo [5].

Figure 1: Exp. 1 Video 2

Figure 2: Exp. 1 Video 3

Figure 3: Exp. 2 Video 4

Figure 4: Exp. 2 Video 9

However, a deep and thorough analysis of user behaviors in the virtual environment still remain undone. How do users explore a new virtual environment? How soon can users find the region of interest (ROI) of this scene? Do users exhibit similar patterns when they explore a virtual environment? The answers to the above questions are yet to be discovered. Understanding the user behavior in virtual environment is beneficial in numerous ways, e.g., improving the usability and utility of the user interface (UI) design for VR applications.

In this paper, we provide a dataset containing the user behavior records captured in a VR spherical video streaming application. As VR streaming is becoming one of the most important VR applications, its differences from the traditional video streaming in user behavior (e.g. users find the ROI in a video by moving their heads) have not been studied yet. Understanding head motions when watching spherical videos can help researchers with various undertakings, for example:

- A better understanding of the user behaviors can facilitate the developments of efficient video transmission system or coding schemes.
- Users exhibit different behavior patterns during spherical video viewing, including head rotation frequency, angular speed, etc, which can help build new user identification mechanism.
- By analyzing users' visual attention, researchers can develop user's attention models in virtual environment. The findings of such research can be used to, for example, improve the UI design of VR applications.

Therefore in this paper we focus on providing a user behavior dataset of spherical video viewing.

MMSys'17, Taipei, Taiwan

DOI: http://dx.doi.org/10.1145/3083187.3083210

Our dataset is collected in two separated experiments in two consecutive weeks. In both experiments, participants are required to watch 9 spherical videos. These videos are selected to represent the most popular video categories of current VR contents. The title sequences and end credits are remove from the original videos, resulting in video durations varying from 2′44″ to 10′55″. 48 participants are involved in the experiments and their head motions are recorded during the viewing, including rotations and positions.

These two experiments aim to serve two different purposes: In the first experiment, we intend to capture the natural user behaviors when immersed in a new virtual environment. As VR devices attract increasing attention, a great number of spherical videos have been made, varying greatly in content, genre and video-recording methods. Thus participants are presented with a set of very different videos, for instance, freestyle skiing video[1] (Fig. 1) or sci-fi video 'Help'[2] (Fig. 2). Each video can be regarded as a unique VR scene and when watching such a video, participants will have similar experience as exploring a virtual environment.

In the second experiment, the target application is live VR streaming. There are two major difference between live VR streaming and on-demand VR streaming, other than its purpose of broadcasting live events. First, the positions and shooting directions of cameras used in live VR streaming are often fixed (e.g. Fig. 3 and Fig. 4). Second, there are often only one ROI in the screen, for example, the stage or the sport field. Despite that live VR streaming may switch between cameras that are in different positions, the single ROI is always located at the center of a video frame. Thus live VR steaming videos are significantly different from other spherical videos. In this experiment, participants are only presented with videos recorded from live events, for instance, a basketball match (Fig. 3) or a talkshow program (Fig. 4). To mimic the real world live streaming scenario, e.g. participants focusing on the video content rather than playing with the VR devices, participants are told they will be taking a short test after the experiment to evaluate their memories of the video contents. In the first experiment, users are also presented with live VR streaming videos to evaluate their behavior differences under different experiment setups.

To the best our knowledge, no existing public dataset contains the user behavior records of spherical video viewing, along with the user demographic profiles. Yu et al.[10] and Corbillon et al.[2] used a small dataset of user behavior during spherical video viewing provided by Jaunt Inc. However, this dataset is not yet public, and both the number of participants and the duration of the videos are much smaller compared to our dataset (eleven people watch eleven 10s long spherical video). The contributions of our dataset lies not only on its size, but also the diversity of participants: Our dataset are collected from participants with various VR experiences level. The participants' demographic records included in the dataset and participants' behavior difference between the first experiment and the second experiment can help research explore the user behavior difference caused by previous experience. Last but not least, our dataset not only provides the head movement records, but also includes the videos used in the experiment.

[1]https://youtu.be/0wC3x_bnnps
[2]https://youtu.be/G-XZhKqQAHU

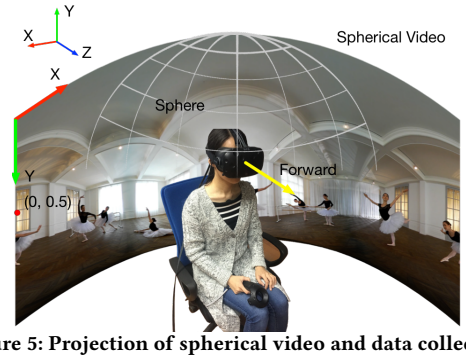

Figure 5: Projection of spherical video and data collection

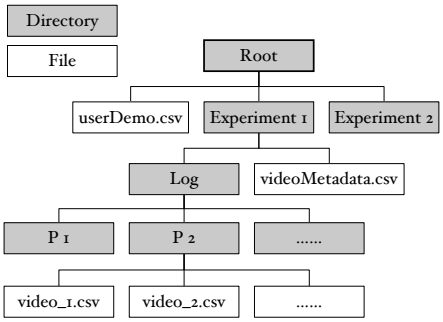

Figure 6: The Directory Structure of the Datasets

The rest of this paper is organized as follows. Section 2 presents the data collection methodology and describes the dataset. Section 3 provides some of our preliminary analyses on our dataset.Section 4 gives the examples of the use of our datasets. Finally, the paper is concluded in section 5.

2 DATASET

2.1 Data Collection Procedure

The VR device used in this paper is a HTC Vive headset. We developed a Unity 3D program which projects equirectangular spherical videos onto the inside of a sphere. Participants are instructed to sit on a 360-degree swivel chair, put on the headset, and interact with the program using a handheld controller, as shown in Fig. 5 After the program is started, participants are exposed to a demo video which helps them to adapt to the virtual reality environment. When participants have acclimatized themselves to the VR environment and are familiar with the operation procedure, they can start the experiment by pressing a button. During the video playback intervals, participants can choose to take a break and start again when they feel comfortable.

In the first experiment, participants are free to look around, same as exploring a virtual reality game scene. Participants are not required to focus on the content of the videos, thus may exhibit their unique behavior patterns or habits. In the second experiment, participants are instructed to focus on the contents of the videos before the experiment. For example, when watching a basketball game, participants can pay attention to how many does each team score. Participants are told that they will take a test after watching all the videos. 10 simple questions in Chinese are randomly selected from a small question bank consisted of 36 questions. This

requirement is made due to the concern that if a user is not interested in the video content, he/she may starts to play with the VR devices or even stops watching. Some sample questions are listed as follows:

- In the 2nd video, how many people are in the band?
- In the 5th video, what color does the winner wear?
- In the 8the video, what color does the singer dye his hair?
- In the 9th video, which guest is a real start-up company founder?

2.2 Availability and Format

The two dataset can be downloaded from our website[3]. The size of our dataset is about 228 MB after compression. Fig. 6 shows the hierarchical structure of our datasets.

For each participant and each video, the following data is stored in csv format (9 fields):

- Timestamp (T): the local time in (UTC/GMT) in "yyyy-MM-dd HH:mm:ss.fff" format
- PlaybackTime (t): the video playback time in seconds
- UnitQuaternion[4] (x, y, z, w): the unit quaternion of the HMD device
- HmdPosition (x, y, z): the position of the HMD device in the Unity 3D world space

2.3 Video Metadata

Table 1: Video Metadata

	No	Len	Content	Category
Experiment 1	1	2′44″	Conan360°-Sandwich	Performance
	2	3′21″	Freestyle Skiing	Sport
	3	4′53″	Google Spotlight-HELP	Film
	4	2′52″	Conan360°-Weird Al	Performance
	5	3′25″	GoPro VR-Tahiti Surf	Sport
	6	10′55″	The Fight for Falluja	Documentary
	7	7′31″	360° Cooking Battle	Performance
	8	2′44″	LOSC Football	Sport
	9	4′52″	The Last of the Rhinos	Documentary
Experiment 2	1	4′38″	Weekly Idol-Dancing	Performance
	2	5′41″	Damai Music Live-Vocie Toy	Performance
	3	3′07″	Rio Olympics VR Interview	Talkshow
	4	6′01″	Female Basketball Match	Sport
	5	2′52″	SHOWTIME Boxing	Sport
	6	3′25″	Gear 360: Anitta	Performance
	7	6′13″	CCTV Spring Festival Gala	Performance
	8	3′46″	VR Concert	Performance
	9	8′40″	Hey VR Interview	Talkshow

In the *videoMetadata.csv* file, we provide a detailed information about the videos that we used in our experiments. 18 videos, as shown in Table 1, are used in these experiments. The metadata stored in the csv file contains the following fields:

- VideoNo: the order that this video this played in the corresponding experiment
- VideoResolution: the resolution of video in the number of pixels (e.g., 1920 × 1080).
- VideoDuration: the temporal length of the video file in seconds.
- FrameRate: the frame rate of the video
- FrameCount: the total number of frames included in the video file
- Bitrate: the bitrate of the video in *kbps*
- VideoLink: the original link to this video
- DropboxLink: the link to download the video clip used in the experiment

2.4 Explanation

Fig. 5 illustrates the left-handed coordinate system adopted by the Unity 3D platform, in which the positive x, y and z axes point right, up and forward, respectively. The unit quaternion (qx, qy, qz, qw) represents the rotations of a objects in 3D space. For each object such as the HMD device in this coordinate system, the forward vector (i.e., a unit vector (x, y, z) represents which direction a participant is looking at) can be calculated from the unit quaternion as follows:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 * qx * qz + 2 * qy * qw, \\ 2 * qy * qz - 2 * qx * qw \\ 1 - 2 * qx^2 - 2 * qy^2 \end{bmatrix}$$

The forward vector is $< 0, 0, 1 >$ if this object has't been rotated, i.e., when its unit quaternion is $< 0, 0, 0, 1 >$. Our program projects the video to the inside of a sphere as texture, starting from the positive x axes. For example, the point $< 0, 0.5 >$ on the video is projected to $< 1, 0, 0 >$ on the sphere, as shown in Fig. 5. The rest of the video is "wrapped" to the sphere anti-clockwise

2.5 Participants

Table 2: Demographic Profile of the Participants

Gender		Age		
Male	Female	≤ 20	20 ∼ 25	≥ 26
24	24	10	31	7

Academic Background		
Undergraduate	Master	PhD
10	36	2

VR Experience			
Never	Heard of before	Sometimes	Used frequently
1	22	23	2

Table 2 presents the demographic profile of all the participants. 48 subjects of different age and gender participated in this experiment. 50% of the participants are male and 64.6% are 20 to 25 years old. 23 participants used VR devices before and two used VR devices frequently. 15 participants have used high-end VR devices like HTC Vive before, while the others have used smartphone VR headsets or standalone VR headsets.

[3] https://wuchlei-thu.github.io/
[4] https://en.wikipedia.org/wiki/Quaternions_and_spatial_rotation

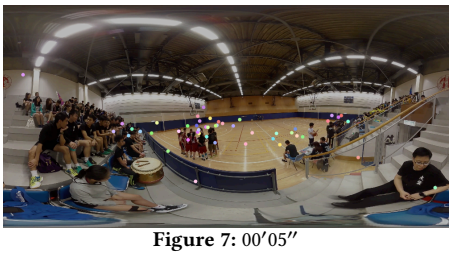
Figure 7: 00′05″

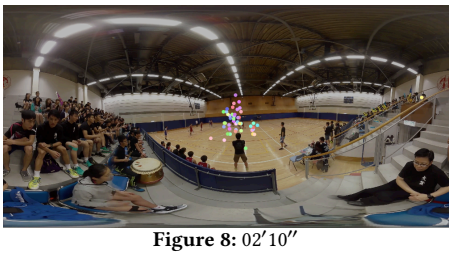
Figure 8: 02′10″

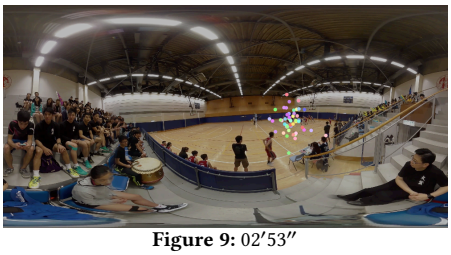
Figure 9: 02′53″

Participants' gazing directions during the viewing the 4th of Experiment 2.

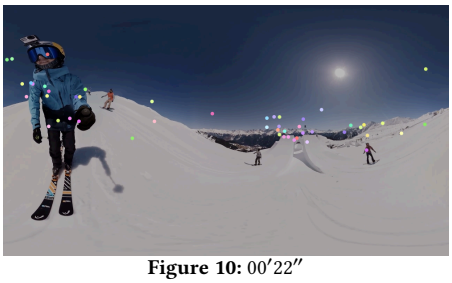
Figure 10: 00′22″

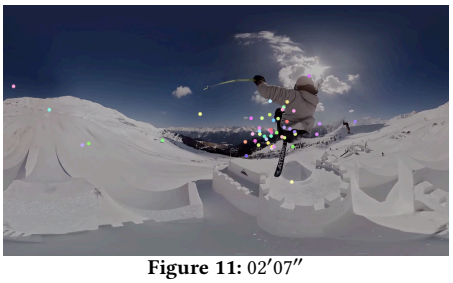
Figure 11: 02′07″


Figure 12: 02′35″

Participants' gazing directions during the viewing the 2nd of Experiment 1.

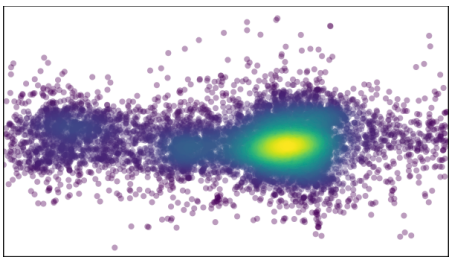
Figure 13: Exp. 1 No. 2

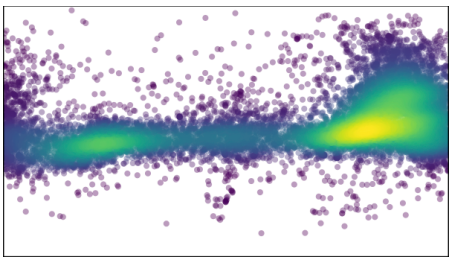
Figure 14: Exp. 1 No. 3

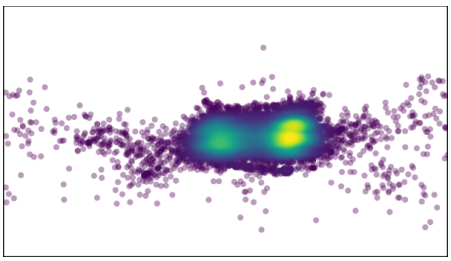
Figure 15: Exp. 2 No. 4

Density maps of participants' gazing directions.

3 DATASET VISUALIZATION AND MEASUREMENT

In this section, we present some of the characteristics and statistics of our datasets. Our program records data every 10ms approximately, i.e., about 100 records every second.

3.1 Visualization

A participant gazing direction, i.e., where the participant is looking, can be projected back onto the video, as shown in Fig. 7 to Fig. 12. Fig. 7 to Fig. 9 show the projections of participants watching a basketball match. In Fig. 7, the video just started and the basketball game hadn't started yet, participants' gazing direction varied greatly from each other. In Fig. 8, the basketball had already started and players from two teams gathered together, the participants' gazing directions also gathered together. In Fig. 9, as players moved from one side of the court to the other, participants' gazing directions moved with them. Fig. 10 to Fig. 12 show the projections of participants observing a freestyle skiing video. We observe that the distributions of gazing points are obviously more sparse compared to the basketball match, since skiers often show up in every directions. Based on these figures, we can conclude that for different videos and different video playback time, the movements of participants' gazing directions are often different.

Fig. 13 to Fig. 15 present some of the density maps of participants' gazing directions during the viewing of three different spherical videos. We can observe that the two density maps of Experiment 1 are more scattered than the one of Experiment 2, and the high-density areas are much larger. This correspond to the facts that the videos selected for Experiment 1 often have multiple ROIs in one frame, for example the heroine and the "monster" in Fig. 2. As the density map Fig. 15 of Experiment 2 indicates, participants are more likely to view content in the vicinity of the equator in Experiment 2, since live VR streaming videos usually have only one ROI which is often located near the equator, for example, the basketball court in Fig. 3 or the talkshow host and guests in Fig. 4.

In Fig. 16 to Fig. 18, we plot the trajectories of participants gazing directions between 60″ and 120″ in 3D space. Fig. 16 and Fig. 17 depict the trajectory of participants' attention during the viewing of the second and third video of Experiment 1, while Fig. 17 is of the fourth video of Experiment 2. We choose the time span [60″, 120″] to avoid the beginning phase of the video session, in which the participants are likely to explore the virtual environments randomly. From these three figures, we can observe that when viewing the videos of Experiment 1, participants have much more diversified gazing direction paths than Experiment 2, for example, in Fig. 17, participants' gazing directions are divided into two groups between 60″ and 70″.

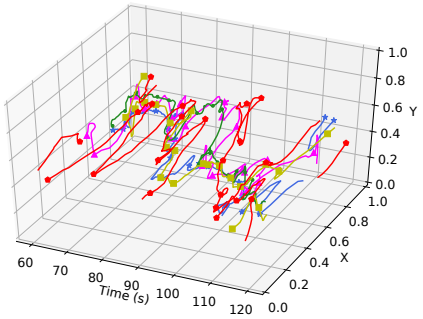

Figure 16: Exp. 1 No. 2

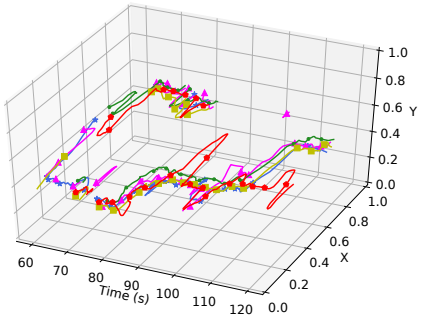

Figure 17: Exp. 1 No. 3

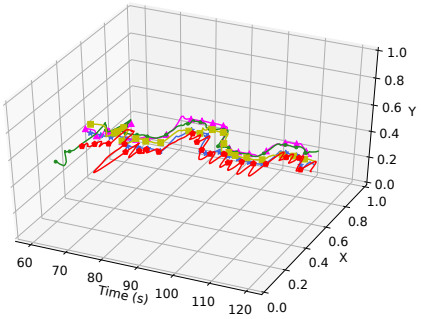

Figure 18: Exp. 2 No. 4

The paths of participants' gazing directions between 60″ and 120″ in 3D space.

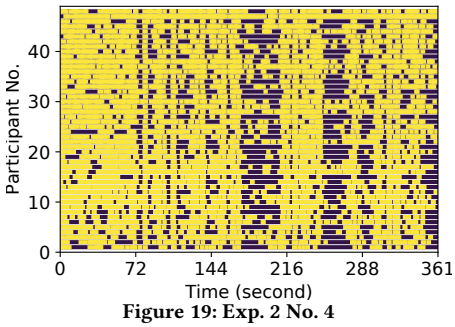

Figure 19: Exp. 2 No. 4

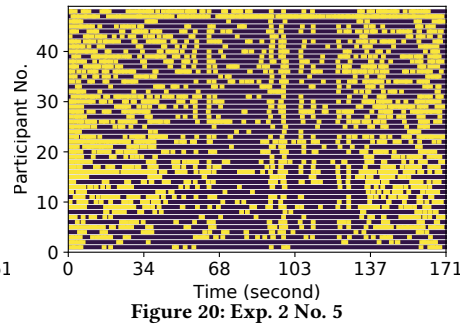

Figure 20: Exp. 2 No. 5

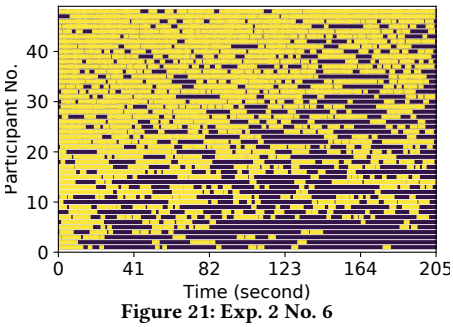

Figure 21: Exp. 2 No. 6

Participants' head movement status during the viewing of three videos in Experiment 2.

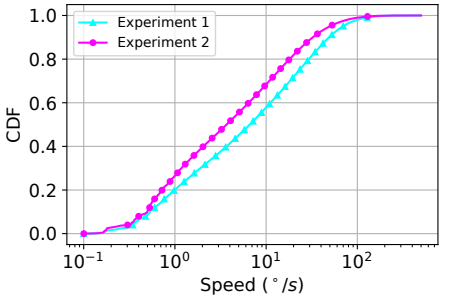

Figure 22: Angular Velocity of Head Motion

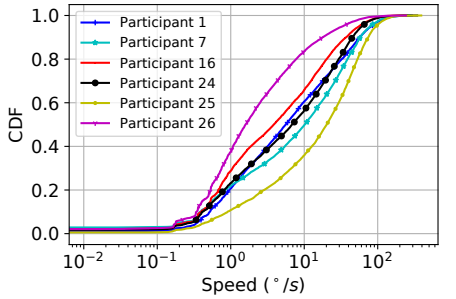

Figure 23: Individual Speed (Experiment 1)

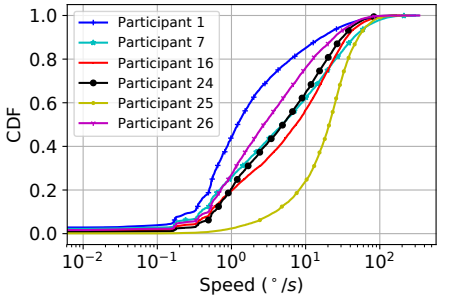

Figure 24: Individual Speed (Experiment 2)

To investigate the head rotation frequencies and amplitudes in live VR streaming, we plot participants' head movement status in Fig. 19 to Fig. 21, in which yellow blocks represents participant has made a large head movement (in this case, a angle larger than 20°) and purple blocks represent participant's head is stationary or the movement is within 20°. From Fig. 19, We can observe that when watching the 4th video in Experiment 2, participants often turn their heads at some time timestamp together (e.g. around 216 second), indicating for some live VR streaming events, there exists some moment that participants tend to make simultaneous head movements. In Fig. 20, we can observe that participants turn their heads less frequently, comparing to Fig. 19, but still tend to turn their heads together at some moments. In Fig. 21, we can observe that participants turn their head much more frequently, however the movements are also more diversified.

3.2 Statistics

Fig. 22 presents the CDF of the angular speed of participants turning their heads during the two experiments, which are sampled three times every second. We can observe that when watching the first set of videos, the angular speed is clearly higher than that in the second experiment. In the first experiment, 57.8% of the speed records are lower than $10°/s$, and 90.5% are lower than $50°/s$. In the second experiment, the number is 68.5% and 95% respectively. We conclude that during the first experiments, participants not only have more diversified regions of interest, but also turn their heads more quickly and frequently. This indicates that for different video sets, the user behavior patterns are different.

Next, we selected 6 participants and calculated their angular speeds during the two experiments. Fig. 23 presents the head angular speed of each individual during the first experiment. We can observe that there exists significantly difference between each participant's head angular speed. For example, the average angular speed of participant 26 is $6.7°/s$, and 83.5% of the speed records

are lower than $10°/s$. While for participants 25, the average speed is $30.5°/s$ and only 35.9% of the speed records are lower than $10°/s$. In Fig. 24, the average speeds of participants 26 and 25 are 8.5 and 25.1, respectively. And the percentage of speed records that are lower than $10°/s$ are 75.9% and 24.1%, respectively. Thus we can conclude that for different users and different video, head rotations as an important user behavior are quite different.

4 APPLICATIONS AND RELATED WORK

Our dataset can be used for a variety of applications. In this section, we provide some of the use cases and the related works.

4.1 Visual Attention Modeling in Virtual Environment

Visual attention modeling as an active research topic[1] is mostly based on eye movement datasets of still images or 2D videos, the visual attention model in virtual environment hasn't been explored. Developing a visual attention model for virtual environment can help improve many VR video or game related applications and techniques, such as ROI adaptive coding. ROI adaptive coding is a popular solution for video streaming over bandwidth-limited networks. Grois et al.[3] proposed to adaptively set the location, size and resolution of the desirable ROI, based on the network condition and configurations, thus providing users with a high-quality decoded video stream, containing the desired ROI. The choices of ROI have a great impact on the video encoding quality, which is often aided by visual attention modeling.

4.2 Tiling based Live VR Streaming with Gazing Prediction

To deal with enormous bandwidth requirement of high resolution live VR streaming, researchers and industries often adopt a tiling based solution, in which viewers will receive a part of the video at a high resolution, and the rest of the video will not be streamed or streamed at a low resolution. For example, a Netherlands company, TNO, adopted a solution to cut the video into multiple tiles and only transfer tiles that are relevant to user's actual viewport[5]. However, switching between tiles causes considerable delays, for instance the system proposed by Ochi et al.[8] takes 2.6 seconds on average to switch to a new tile. This weakness significantly spoiled the immersive experience. Using our dataset can facilitate the research and development of live VR streaming system. For example, if the live VR streaming can detect if a user is about to turn his/her head based on the data collected before, then it's possible to prepare and fetch the tiles that may be used in advance.

4.3 User Identification Based on Head Motion Pattern

As the virtual reality and augmented reality (AR) devices getting increasingly popular, the security concern with such devices is also receiving notable attentions. Previous works have proved that data collected by sensors (e.g. accelerometer) can be leveraged to identify user. For example, Mantyjarvi et al.[7] proposed to identify users using gait signals collected from accelerometers of wearable devices on users' belts. Hallac et al.[4] proposed to identify drivers from sensor data collected at a single turn and demonstrated that turns are suitable for drivers identification compared to driving straightaway. As our datasets contain the fine-grained information of user's head movements, it's plausible to identify a user out of the others. An related application may be developing new VR application login authentication methods based on user identification.

[5]https://www.tno.nl/en/about-tno/news/2016/8/ibc-2016-sees-industry-s-first-ultra-high-quality-virtual-reality-streaming-solution/

5 CONCLUSION

In order to facilitate the development of VR systems and applications, we present a user behavior dataset of user viewing spherical videos. This dataset can be used to improve the design of various spherical video related applications, such as live VR streaming system. Besides these applications, this dataset can also be used to explore new user identification mechanism using VR devices, visual attention model in virtual environment, etc.

ACKNOWLEDGEMENT

This work is supported in part by the National Natural Science Foundation of China under Grant No. 61402247 and U1611461.

REFERENCES

[1] A. Borji and L. Itti. 2013. State-of-the-Art in Visual Attention Modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 1 (Jan 2013), 185–207. DOI: https://doi.org/10.1109/TPAMI.2012.89

[2] Xavier Corbillon, Alisa Devlic, Gwendal Simon, and Jacob Chakareski. 2016. Viewport-Adaptive Navigable 360-Degree Video Delivery. *arXiv preprint arXiv:1609.08042* (2016).

[3] D. Grois, E. Kaminsky, and O. Hadar. 2010. ROI adaptive scalable video coding for limited bandwidth wireless networks. In *2010 IFIP Wireless Days*. 1–5. DOI: https://doi.org/10.1109/WD.2010.5657709

[4] D. Hallac, A. Sharang, R. Stahlmann, A. Lamprecht, M. Huber, M. Roehder, R. Sosič, and J. Leskovec. 2016. Driver identification using automobile sensor data from a single turn. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. 953–958. DOI: https://doi.org/10.1109/ITSC.2016.7795670

[5] G. A. Koulieris, G. Drettakis, D. Cunningham, and K. Mania. 2016. Gaze prediction using machine learning for dynamic stereo manipulation in games. In *2016 IEEE Virtual Reality (VR)*. 113–120. DOI: https://doi.org/10.1109/VR.2016.7504694

[6] M. Lee, K. Kim, S. Daher, A. Raij, R. Schubert, J. Bailenson, and G. Welch. 2016. The wobbly table: Increased social presence via subtle incidental movement of a real-virtual table. In *2016 IEEE Virtual Reality (VR)*. 11–17. DOI: https://doi.org/10.1109/VR.2016.7504683

[7] J. Mantyjarvi, M. Lindholm, E. Vildjiounaite, S. M. Makela, and H. A. Ailisto. 2005. Identifying users of portable devices from gait pattern with accelerometers. In *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, Vol. 2. ii/973–ii/976 Vol. 2. DOI: https://doi.org/10.1109/ICASSP.2005.1415569

[8] Daisuke Ochi, Yutaka Kunita, Kensaku Fujii, Akira Kojima, Shinnosuke Iwaki, and Junichi Hirose. 2014. HMD viewing spherical video streaming system. In *Proceedings of the ACM International Conference on Multimedia*. ACM, 763–764.

[9] A. Steed, S. Frlston, M. M. Lopez, J. Drummond, Y. Pan, and D. Swapp. 2016. An 'In the Wild' Experiment on Presence and Embodiment using Consumer Virtual Reality Equipment. *IEEE Transactions on Visualization and Computer Graphics* 22, 4 (April 2016), 1406–1414. DOI: https://doi.org/10.1109/TVCG.2016.2518135

[10] M. Yu, H. Lakshman, and B. Girod. 2015. A Framework to Evaluate Omnidirectional Video Coding Schemes. In *2015 IEEE International Symposium on Mixed and Augmented Reality*. 31–36. DOI: https://doi.org/10.1109/ISMAR.2015.12