A Dataset of Head and Eye Movements for 360° Videos

Erwan J. David
LS2N UMR CNRS 6004
Université de Nantes
Nantes, France
erwan.david@univ-nantes.fr

Jesús Gutiérrez
LS2N UMR CNRS 6004
Université de Nantes
Nantes, France
jesus.gutierrez@univ-nantes.fr

Antoine Coutrot
LS2N UMR CNRS 6004
Université de Nantes
Nantes, France
antoine.coutrot@ls2n.fr

Matthieu Perreira Da Silva
LS2N UMR CNRS 6004
Université de Nantes
Nantes, France
matthieu.perreiradasilva@
univ-nantes.fr

Patrick Le Callet
LS2N UMR CNRS 6004
Université de Nantes
Nantes, France
patrick.lecallet@univ-nantes.fr

ABSTRACT

Research on visual attention in 360° content is crucial to understand how people perceive and interact with this immersive type of content and to develop efficient techniques for processing, encoding, delivering and rendering. And also to offer a high quality of experience to end users. The availability of public datasets is essential to support and facilitate research activities of the community. Recently, some studies have been presented analyzing exploration behaviors of people watching 360° videos, and a few datasets have been published. However, the majority of these works only consider head movements as proxy for gaze data, despite the importance of eye movements in the exploration of omnidirectional content. Thus, this paper presents a novel dataset of 360° videos with associated eye and head movement data, which is a follow-up to our previous dataset for still images [14]. Head and eye tracking data was obtained from 57 participants during a free-viewing experiment with 19 videos. In addition, guidelines on how to obtain saliency maps and scanpaths from raw data are provided. Also, some statistics related to exploration behaviors are presented, such as the impact of the longitudinal starting position when watching omnidirectional videos was investigated in this test. This dataset and its associated code are made publicly available to support research on visual attention for 360° content.

CCS CONCEPTS

• **Information systems** → **Multimedia databases**; • **Computing methodologies** → Virtual reality;

MMSys'18, June 12–15, 2018, Amsterdam, Netherlands

ACM ISBN 978-1-4503-5192-8/18/06…$15.00
https://doi.org/10.1145/3204949.3208139

KEYWORDS

Omnidirectional video, 360° videos, dataset, eye-tracking, saliency, gaze behavior

ACM Reference Format:
Erwan J. David, Jesús Gutiérrez, Antoine Coutrot, Matthieu Perreira Da Silva, and Patrick Le Callet. 2018. A Dataset of Head and Eye Movements for 360° Videos. In *MMSys'18: 9th ACM Multimedia Systems Conference, June 12–15, 2018, Amsterdam, Netherlands.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3204949.3208139

1 INTRODUCTION

Virtual Reality and 360° content (among other emerging immersive media technologies, such as augmented reality or light field imaging) is providing users with new interactive experiences and more freedom to explore the represented scenes. These new possibilities entail a great change on how users interact with media technologies, since they can look wherever they want and are no longer limited, as in traditional media, to passively look at what they are shown.

In this sense, these novelties should be taken into account to design and develop appropriate immersive media systems, such as efficient encoding, transmission and rendering techniques to provide the best quality to the end users. With this aim, understanding how people observe and explore 360° is crucial.

In fact, some works have already been presented dealing with the study of visual attention in VR and 360° content. For example, a preliminary study was carried out by Marmitt and Duchowski [11] analyzing head and eye movements to investigate visual scanpaths in VR environments. This work has been recently picked up taking advantage of new and improved VR devices to analyze exploring behaviors of users when watching 360° images, recording head and eye movements with eye-trackers embedded in VR headsets [14][20]. While eye movements analysis has proved to provide an important added value to visual attention modeling in VR [16], gaze data is not always easily accessible. Thus, head movements could be considered as a valuable proxy [24][22]. Studies have been presented analyzing head movements during 360° images exploration [2][5][20].

The way people watch 360° images may substantially differ from how they explore omnidirectional videos, where their attention can

be more guided by the dynamic content. Therefore, some studies have already focused on analyzing visual attention in 360° dynamic content. For instance, Serrano *et al.* used head and eye tracking data for movie editing and segmentation [19]. Nevertheless, the aforementioned difficulty to gather eye tracking data has caused the majority of studies in this topic to only consider head movements. For example, Su *et al.* [21] used head movements for automatic cinematography in 360° videos. Also, Corbillon *et al.* studied the application of head movements for efficient delivery of 360° video [3]. Similarly, Wu *et al.* analyzed exploring behaviors based on head tracking data for video streaming of omnidirectional video [23], also releasing a dataset.

The importance of public datasets containing stimuli and eye-tracking information is crucial for the research community to evolve in the development of efficient techniques for coding, transmitting, and rendering 360° content. A good example was the publication of the dataset of head and eye movements for omnidirectional images by Rai *et al.* [14], for the "Salient360!" Grand Challenge at ICME'17 promoting the research on models for saliency and scanpath prediction for 360° images, and resulting in the publication of several valuable models [17]. In addition, other datasets have been published recently with 360° videos and solely head movement data, such as the 360° video head movement dataset from Corbillon *et al.* [2] (five videos from *Youtube*, duration of 70 seconds, watched by 59 users), the 360° video viewing dataset in head-mounted VR, by Lo *et al.* [10] (ten videos from *Youtube*, duration of one minute, watched by 50 users), and the dataset for emotion induction research by Li *et al.* [9] containing head movement data and corresponding ratings of arousal and valence (73 videos from *Youtube*, durations from 29 to 668 seconds, watched by 95 users).

Taking this into account, and complementing previous works on datasets of head movements in video and head and eye movements in still images, this paper presents a dataset of videos containing head and eye tracking data for research on exploring behaviors with 360° dynamic content. The dataset contains 19 videos in equirectangular format with associated data of head and eye movements collected from a free-viewing experiment with 57 observers wearing a VR headset with an integrated eye-tracker. In addition, we analyze users' exploring behavior such as the impact of starting longitudinal positions on 360° content exploration, and the relation between head and eye movement data.

The rest of the paper is organized as follows. Section 2 presents the subjective experiment carried out to create the datasets, as well as the details about gathered data. Then, Section 3, describes how raw gaze data obtained from the subjective experiment were processed to generate visual attention data. Section 4 presents statistical results related to the exploration of 360° content. Finally, some conclusions are provided in Section 5.

2 DATASET & SUBJECTIVE EXPERIMENT

2.1 Video stimuli

This dataset is composed of 19 videos gathered from *Youtube* (see supplementary material for frame examples of each video). All videos are 4K in resolution (3840x1920 pixels), equirectangular format, their main properties are shown in Table 1. In particular, the category indicates some high-level attributes (e.g, indoor/outdoor, rural/natural, containing people faces, etc.). In addition, the Spatial perceptual Information (SI) and the Temporal perceptual Information (TI) [6] were computed for all videos in equirectangular format (using an SI filter of 13x13 pixels [13]). In addition to the objective of covering a wide range of these features, these videos were also selected taking into account their license of use (Creative Commons), and their duration with uninterrupted content (no camera cuts). Specifically, a duration, albeit short, of 20 seconds was considered to abide by these last two constraints.

2.2 Equipment

360° videos were displayed in a VR headset (HTC VIVE, HTC, Valve corporation) equipped with an SMI eye-tracker (*SensoMotoric Instrument*). The HTC VIVE headset allows sampling of scenes by approximately 110° horizontal by 110° vertical field of view (1080x1200 pixels) at 90 frames per second. The eye-tracker samples gaze data at 250Hz with a precision of 0.2°. A custom *Unity3D* (Unity Engine, CA, USA) scene was created to display videos. Equirectangular content was projected onto a virtual sphere via a shader program computing an equirectangular-to-sphere projection on the GPU. A process independent from the Unity Engine process was used to write HMD and eye-tracker data to disk at the speed of the eye-tracker sampling rate. The experiment was running on a computer with an NVIDIA GTX1080 GPU.

2.3 Observers

57 participants were recruited (25 women; age 19 to 44, mean: 25.7 years), normal or corrected-to-normal vision was verified with the Monoyer test, acceptable color perception was tested with the Ishihara test. Dominant eye of all observers was checked. Participants received monetary compensation for their time. All 19 videos were observed by all observers for their entire duration (20 seconds).

2.4 Viewing procedure

Observers were told to freely explore 360° videos as naturally as possible while wearing a VR headset. Videos were played without audio.

In order to let participants safely explore the full 360° field of view, we chose to have them seat in a rolling chair. The fact that participants are not aware of their surroundings while wearing a HMD is hazardous (e.g. colliding with furnitures, falling over). Additionally, the HMD's cable is an inconvenience when standing and exploring a 360° scene.

To study the impact of starting longitudinal positions on content exploration, we added a between-subjects condition where participants could start exploring omnidirectional contents either from an implicit longitudinal center (0° and center of the equirectangular projection) or from the opposite longitude (180°). Videos were observed in both rotation modalities by at least 28 participants each. We controlled observers starting longitudinal position in the scene by offsetting the content longitudinal position at stimuli onset, making sure participants start exploring 360° scenes at exactly 0°, or 180° of longitude according to the modality. Video order and starting position modalities were cross-randomized for all participants.

Table 1: Main properties of the omnidirectional video dataset.

Title	Frame-rate	Category	Camera traveling	SI	TI
Abbottsford	30	Indoor, Urban, People	No	84.336	1.422
Bar	25	Indoor, Urban, People	Yes	119.810	27.578
Cockpit	25	Indoor, Urban, People	Yes	53.487	26.371
Cows	24	Outdoor, Rural	No	48.125	2.059
Diner	30	Indoor, Urban, People	No	60.663	2.425
DroneFlight	25	Outdoor, Urban	No	48.350	13.254
GazaFishermen	25	Outdoor, Urban, People	No	92.732	1.246
Fountain	30	Outdoor, Urban	No	67.346	14.665
MattSwift	30	Indoor, Urban, People	No	84.675	3.361
Ocean	30	Outdoor, Water, People	No	25.885	11.103
PlanEnergyBioLab	25	Indoor, Urban, People	No	65.181	4.012
PortoRiverside	25	Outdoor, Urban, People	No	52.655	3.201
Sofa	24	Indoor, Urban, People	No	83.546	1.069
Touvet	30	Outdoor, Urban	Yes	59.520	4.897
Turtle	30	Outdoor, Rural, People	No	32.351	9.531
TeatroRegioTorino	30	Indoor, Urban, People	No	63.064	5.983
UnderwaterPark	30	Outdoor, Natural	Yes	42.082	9.793
Warship	25	Indoor, Urban, People	No	49.939	5.359
Waterpark	30	Outdoor, Urban, People	Yes	57.625	27.022

Observers started the experimentation by an eye-tracker calibration, repeated every 5 videos to make sure that eye-tracker's accuracy does not degrade. the total duration of the test was less than 20 minutes.

2.5 Dataset structure

Organization of the dataset into folder is illustrated in Fig. 1. The video dataset is found in the "Stimuli" folder, arranged in no particular order.

Visual attention data is organized in folders according to their data type: whether if they come from Head-only (**H**) or head and eye movements (**H+E**). For both cases, saliency maps are stored in a folder named "SalMaps". It contains one saliency map per stimulus as a compressed binary file; filenames provides information under the following convention: *title_WxHxFc_Enc.tar.gz* where *title* is the stimuli name, H and W saliency map's height and width in pixels, *Fc* is the frame count, *Enc* is float precision. A python script named *readBinarySalmap.py* is provided as an example on how to read uncompressed saliency map binary files.

The *Scanpaths* directory contains a CSV text file for each stimulus. For H+E data, scanpaths from left and right gaze are provided independently. CSV files contain all identified fixations for one video, ordered temporally for each observer one after the other in the file. The first data column reports fixation indexes for each participants, this value is incremented with each new fixation until reaching the end of an observer's trial, after which indexing starts over at 0 for the next observer. Next two columns are gaze positions in longitudes and latitudes, normalized between 0 and 1; longitudes should be multiplied by 2π and latitudes by π to obtain positions on the sphere. To display fixation positions in an equirectangular map, multiply the same normalized longitudes and latitudes respectively by the desired width and height of the equirectangular map. Next two columns encode starting timestamp and duration of fixations (in msec.). Finally, the last two columns report fixations start and end frames (integers).

Last directory, *Tools*, contains python scripts. *saliencyMeasures.py* (adapted from the MIT Saliency Benchmark matlab toolbox [1]s) and *scanpathMeasure.py* contains saliency and scanpaths similarity measure implementations as well as examples of their use. *readBinarySalmap.py* explains with an example how frames from binary saliency map are to be extracted.

3 GAZE PROCESSING

In order to obtain fixations necessary for the creation of saliency maps and scanpath files we process raw gaze data from the eye-tracker system.

Data acquired from the system is sampled every 4 msec. (eye-tracker's sampling rate). Each sample contains the following information: camera rotation (camera Euler angles as proxy for the HMD/Head rotation); information about left, right and mean gaze direction as a unit vector (3D vector in world space) relative to the camera rotation; left and right cameras baseline (vector distance from central camera to "eye" cameras in world space); left and right eyes 2D gaze positions mapped onto a virtual viewport (2160x1200 pixels, 111.9x105.6 degrees). To project raw viewport data onto a unit sphere, we use camera rotation data to create rotation matrices R (equation 1) which are multiplied with 2D gaze data transformed into a 3D vectors g (equation 2), the resulting 3D vectors are subsequently normalized.

$$R = \begin{pmatrix} \cos\phi\cos\theta & \cos\phi\sin\theta\sin\psi - \sin\phi\cos\psi & \cos\phi\sin\theta\cos\psi + \sin\phi\sin\psi \\ \sin\phi\cos\theta & \sin\phi\sin\theta\sin\psi + \cos\phi\cos\psi & \sin\phi\sin\theta\cos\psi - \cos\phi\sin\psi \\ -\sin\theta & \cos\theta\sin\psi & \cos\theta\cos\psi \end{pmatrix} \quad (1)$$

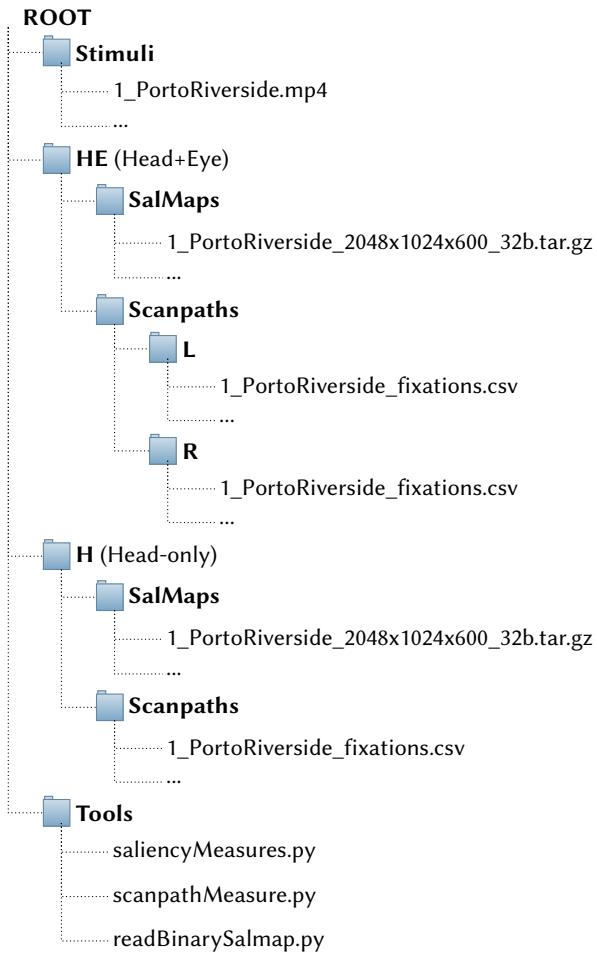

Figure 1: Folder tree composition of the new dataset of 360° videos and head and eye movement data

Where θ is the pitch axis (elevation), ψ the yaw axis (azimuth or heading), ϕ the roll axis (bank).

$$g = \begin{pmatrix} F \\ pxSizeHoriz(P_x - W/2) \\ pxSizeVert(P_y - H/2) \end{pmatrix} \quad (2)$$

Where F is the virtual focal distance (1.5 in our case), $pxSizeHoriz$ the pixel size of one degree horizontally, $pxSizeVert$ the pixel size of one degree vertically, P_x and P_y gaze positions in viewport space, W and H viewport's width and height (pixels).

Four different types of visual attention data are considered. The first to types concern the use of head and eye movements, here gaze and HMD data are processed together to produce gaze positions on the spherical scene. Data is parsed into fixations and saccades, in particular to extract fixations which are periods of reduced eye movements during which scene perception is implied and output saliency maps and scanpath files..

Next two types are based on head-only movements. Here we are purposefully processing head rotation without gaze data and output saliency maps and scanpath files as well. We chose not to define "head saccades" as period of low head rotation velocity [12][20][4][5] since this results in a far reduced number of saccades which do not appear to be an accurate representation of the latent scene perception. Moreover, head movements don't necessarily mean a loss of perception as it is mostly the case during actual saccades [8]; it is the addition of head and eye movements data which will inform us of the actual perception which can be mediated by compensations behaviors between head and eyes. For these reasons, such concepts as head "fixation" and "saccade" are questionable and we resort to head trajectories (subsection 3.2).

3.1 Parsing gaze data to fixations

We rely on a velocity-based algorithm [18] to identify fixations and saccades from eye movements. Because our data is located on a unit sphere we cannot rely on the euclidean distance, as it would mean measuring a line between two points through the sphere; though as sampling rate increases gaze points get closer together spatially and the euclidean distance becomes a good approximation. In spite of this last remark, we define velocity as the orthodromic distance (i.e. great-circle distance, equation 3 shows the Haversine variant used) between two gaze samples divided by their time difference. The 1D velocity signal was smoothed with a gaussian filter ($\sigma = 1$ sample). Gaze samples with velocities below 80°/sec. threshold were categorized as fixations. In a second step, we elected to remove fixations lasting less than 80ms.

$$\Delta\sigma = 2\arcsin\sqrt{\sin^2\left(\frac{\Delta\phi}{2}\right) + \cos\phi_1 \cdot \cos\phi_2 \cdot \sin^2\left(\frac{\Delta\lambda}{2}\right)} \quad (3)$$

Where $\Delta\sigma$ is the distance in angle between two points on the sphere, $\Delta\phi$ the difference in latitudes, $\Delta\lambda$ the difference in longitudes, $\phi 1$ and $\phi 2$ latitudes of the two points compared.

3.2 Head trajectory

Because perception is possible during head movements we chose to model head data as trajectories on the unit sphere. To achieve this we down-sampled the 20 seconds raw gaze data into 100 samples by selecting sequential windows of 200 msec. and computing gaze position centroids on samples within said time windows. One benefit of this method is to obtain data samples aligned between observers for each stimuli, thus settling the issue of scanpath comparison measures usually requiring a method of aligning one scanpath with another. In the case of dynamic contents, familiar methods of alignment (e.g. [7]) imply comparing as peers samples from two different timestamps, thus comparing together samples which occurred during frames displaying different contents.

3.3 Scanpaths

For each video and for both types of data, were extracted sequences of gaze positions on the spherical scene reported as scanpath in the form of text files (described in subsection 2.5). The Head-and-Eye data "Scanpaths" folder contains one folder for each eyes as lateralized scanpaths are provided for analysis.

3.4 Saliency maps

Saliency maps are computed by convolving each fixation or trajectory points (for all observers of one video) with a Gaussian. A sigma of 2° is chosen for head and eye data and 3.34° for head-only data. The former accounts for eye-tracker's precision and foveal perception, the latter is selected in order to stay consistent with the image dataset [14]. This convolution operation is done in 2D sphere space (latitude and longitude coordinates), because an isotropic Gaussian on an equirectangular map would be anisotropic back-projected onto a sphere (see examples of videos saliency in supplementary material). A kernel modeled by a Kent distribution can be used as well but we considered a Gaussian to be an acceptable approximation.

4 RESULTS

We describe below three sets of analyses possible with the data and tools provided. First, fixation count per latitudes and per longitudes. Second analysis shows the similarity between saliency maps obtained from Head-and-Eye data and Head-only data. Finally, we show the similarity between starting rotation conditions as a function of time.

When comparing saliency maps that are equirectangular projections, it is necessary to correct for the latitudinal distortions overrepresenting (in number of pixels) areas closer to the poles. We decided to correct for these distortions by weighting down areas near the poles with a sine function according to the latitude ($\sin y$ for $y \in [0, \pi]$); we found it to be a simpler and more accurate method than a quasi-uniform sampling on a sphere as in [14].

Our data also allows analysis of bottom-up and top-down time-ranges (as described in [15]), it is possible to extract from scanpath CSV files all fixations that occurred in the first 500 msec. of observation, for instance.

4.1 360° content exploration

We report the distribution of fixations as a function of longitudes and latitudes in Fig. 2. Participants observe longitudinally (horizontally on the equirectangular projection) with two peaks arising at 0° and 180°, the two starting rotation modalities. The 0° peak is greater and can be explained by visual stimuli often displaying a center bias even in 360° conditions. Latitudinally, observers are much more inclined to explore areas at and around the equator as we can see by fixation numbers decreasing as a function of distance to the equator (90°).

4.2 Head-Eye and Head-only saliency maps comparison

Head-only and Head-and-eye saliency maps are compared together. To make such comparisons we pooled saliency frames by intervals of 200ms (between 4 and 6 frames according to video frame-rate) which we added together then normalized (divided by the total sum) to obtain new saliency maps for each time increment. For each stimulus we computed a similarity values by computing KLD (Kullback-Leibler Divergence) and CC (Cross-Correlation) for each such saliency maps paired according to the temporal (frame) alignment. The resulting sequence of similarity measures are then averaged over group of frames to obtain a single comparison value as reported in Fig. 3 for each stimulus.

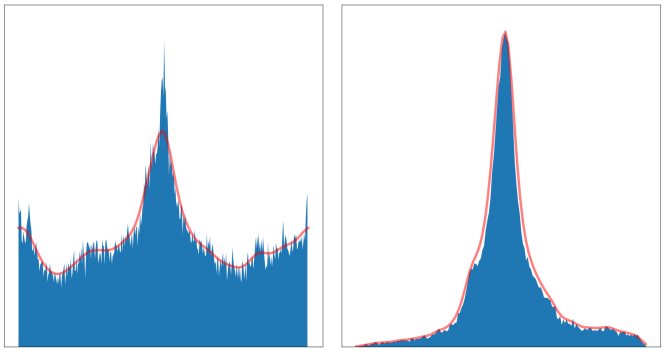

Figure 2: Number of fixations (blue) by longitudes (left, -180° to 180°) and latitudes (right, 0° to 180°). PDF curves (red) are fitted with a von Mises kernel for longitudes and a Gaussian kernel for latitudes.

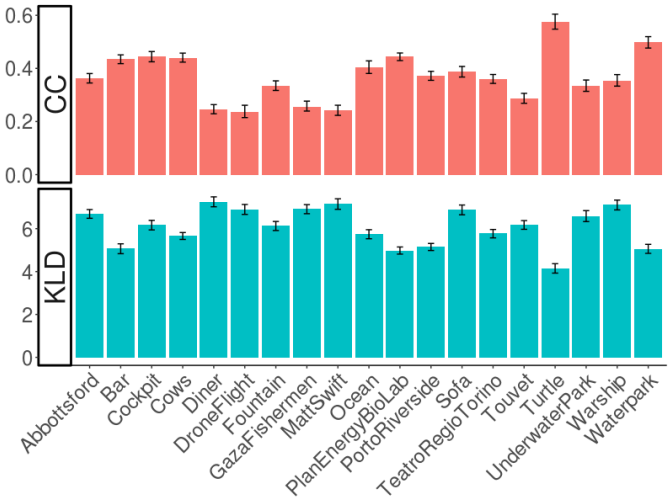

Figure 3: Head-only and Head-and-Eye saliency maps similarity measures: CC (red) and KLD (blue). Error bars report confidence intervals (95%).

The differences observed account for the information lost in Head-only saliency maps relative to Head-and-Eye maps, which is simplistically modeled by a larger Gaussian sigma during the creation of saliency maps in this dataset.

4.3 Starting rotation effect

We compare saliency maps pooled by batches of frames as described in subsection 4.2. Instead of averaging over groups of frames, each saliency map is considered aligned temporally and compared with a Head-and-Eye saliency map according to starting rotation modalities. Averaging over stimuli, we obtain similarity measures (KLD and CC) per frame groups (Fig. 4). Results show that saliency maps are quite dissimilar in the first seconds of exploration. Though, this difference decreases with time and shows no improvements after approximately 5 seconds of exploration.

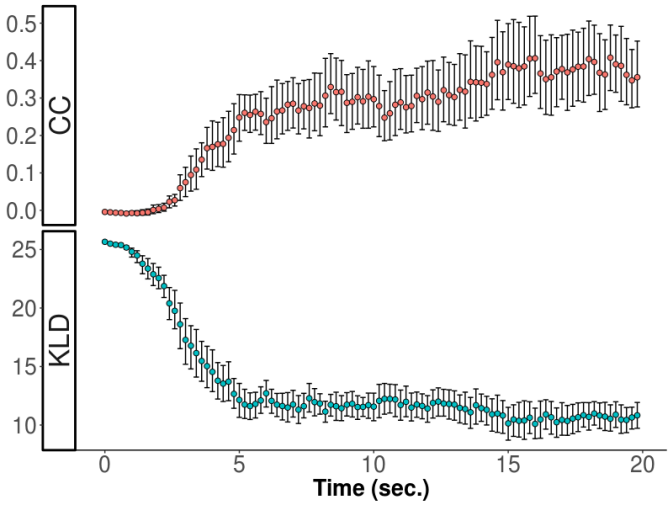

Figure 4: Saliency maps similarities between starting rotation condition groups as a function of time. Similarity measures reported: CC (red) and KLD (blue). Error bars report confidence intervals (95%).

5 CONCLUSION

In this paper a new dataset of 19 omnidirectional videos in equirectangular format is presented, with the associated gaze fixation and head trajectory data in the form of saliency maps and scanpaths. This data was obtained after processing the raw eye and head movements gathered from a free-viewing experiment with 57 observers wearing a VR headset equipped with an eye-tracker. A between-subject condition was added to study the impact of starting longitudinal positions on 360° content exploration. In order to aid the community in analyzing this data, some useful tools to compare saliency maps and scanpaths are also provided.

This dataset extends our previous work with still images [14], providing a public dataset of 360° videos with the added value of eye gaze data, in addition to head movement information, which may help in the research on visual attention in VR and its applications to coding, transmission, rendering and quality assessment of 360° content.

People interested in the dataset can visit *salient360.ls2n.fr/datasets/* in order to access the repository.

ACKNOWLEDGMENTS

The work of Jesús Gutiérrez has been supported by the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement N. PCOFUND-GA-2013-609102, through the PRESTIGE Programme coordinated by Campus France. The work of Erwan David has been supported by RFI Atlanstic2020.

REFERENCES

[1] Zoya Bylinskii, Tilke Judd, Aude Oliva, Antonio Torralba, and Frédo Durand. 2016. What do different evaluation metrics tell us about saliency models? *arXiv preprint arXiv:1604.03605* (2016).

[2] Xavier Corbillon, Francesca De Simone, and Gwendal Simon. 2017. 360-Degree Video Head Movement Dataset. *Proceedings of the 8th ACM on Multimedia Systems Conference - MMSys'17* (June 2017), 199–204.

[3] Xavier Corbillon, Gwendal Simon, Alisa Devlic, and Jacob Chakareski. 2017. Viewport-adaptive navigable 360-degree video delivery. *IEEE International Conference on Communications* (2017).

[4] Yu Fang, Ryoichi Nakashima, Kazumichi Matsumiya, Ichiro Kuriki, and Satoshi Shioiri. 2015. Eye-head coordination for visual cognitive processing. *PloS one* 10, 3 (2015), e0121035.

[5] Brian Hu, Ishmael Johnson-Bey, Mansi Sharma, and Ernst Niebur. 2017. Head movements during visual exploration of natural images in virtual reality. In *2017 51st Annual Conference on Information Sciences and Systems (CISS)*. 1–6. https://doi.org/10.1109/CISS.2017.7926138

[6] ITU. 2008. Subjective video quality assessment methods for multimedia applications. (April 2008).

[7] Halszka Jarodzka, Kenneth Holmqvist, and Marcus Nyström. 2010. A vector-based, multidimensional scanpath similarity measure. In *Proceedings of the 2010 symposium on eye-tracking research & applications*. ACM, 211–218.

[8] Eileen Kowler. 2011. Eye movements: The past 25 years. *Vision research* 51, 13 (2011), 1457–1483.

[9] Benjamin J. Li, Jeremy N. Bailenson, Adam Pines, Walter J. Greenleaf, and Leanne M. Williams. 2017. A Public Database of Immersive VR Videos with Corresponding Ratings of Arousal, Valence, and Correlations between Head Movements and Self Report Measures. *Frontiers in Psychology* 8, DEC (Dec. 2017). https://doi.org/10.3389/fpsyg.2017.02116

[10] Wen-Chih Lo, Ching-Ling Fan, Jean Lee, Chun-Ying Huang, Kuan-Ta Chen, and Cheng-Hsin Hsu. 2017. 360Âř Video Viewing Dataset in Head-Mounted Virtual Reality. *Proceedings of the 8th ACM on Multimedia Systems Conference - MMSys'17* (June 2017), 211–216.

[11] G. Marmitt and A.T. T Duchowski. 2002. Modeling visual attention in vr: Measuring the accuracy of predicted scanpaths. In *Eurographics 2002, Short Presentations*. Saarbrücken, Germany, 217–226. https://doi.org/10.2312/egs.20021022

[12] Gerd Marmitt and Andrew T Duchowski. 2002. *Modeling visual attention in VR: Measuring the accuracy of predicted scanpaths*. Ph.D. Dissertation. Clemson University.

[13] Margaret H. Pinson, Lark Kwon Choi, and Alan Conrad Bovik. 2014. Temporal Video Quality Model Accounting for Variable Frame Delay Distortions. *IEEE Transactions on Broadcasting* 60, 4 (Dec. 2014), 637–649. https://doi.org/10.1109/TBC.2014.2365260

[14] Yashas Rai, Jesús Gutiérrez, and Patrick Le Callet. 2017. A dataset of head and eye movements for 360 degree images. In *Proceedings of the 8th ACM Multimedia Systems Conference, MMSys 2017*. https://doi.org/10.1145/3083187.3083218

[15] Yashas Rai, Patrick Le Callet, and Gene Cheung. 2016. Quantifying the relation between perceived interest and visual salience during free viewing using trellis based optimization. In *Image, Video, and Multidimensional Signal Processing Workshop (IVMSP), 2016 IEEE 12th*. IEEE, 1–5.

[16] Yashas Rai, Patrick Le Callet, and Philippe Guillotel. 2017. Which saliency weighting for omni directional image quality assessment?. In *Quality of Multimedia Experience (QoMEX), 2017 Ninth International Conference on*. IEEE, 1–6.

[17] Salient360. 2018. Special Issue. *Signal Processing: Image Communication* (2018). To appear.

[18] Dario D Salvucci and Joseph H Goldberg. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*. ACM, 71–78.

[19] Ana Serrano, Vincent Sitzmann, Jaime Ruiz-Borau, Gordon Wetzstein, Diego Gutierrez, and Belen Masia. 2017. Movie editing and cognitive event segmentation in virtual reality video. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 47.

[20] Vincent Sitzmann, Ana Serrano, Amy Pavel, Maneesh Agrawala, Diego Gutierrez, Belen Masia, and Gordon Wetzstein. 2018. Saliency in VR: How do people explore virtual environments? *IEEE Transactions on Visualization and Computer Graphics* (2018), 1–1. https://doi.org/10.1109/TVCG.2018.2793599

[21] Yu-Chuan Su, Dinesh Jayaraman, and Kristen Grauman. 2016. Pano2Vid: Automatic Cinematography for Watching 360° Videos. 1 (Dec. 2016).

[22] Evgeniy Upenik and Touradj Ebrahimi. 2017. A simple method to obtain visual attention data in head mounted virtual reality. In *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. 73–78.

[23] Chenglei Wu, Zhihao Tan, Zhi Wang, and Shiqiang Yang. 2017. A Dataset for Exploring User Behaviors in VR Spherical Video Streaming. *Proceedings of the 8th ACM on Multimedia Systems Conference - MMSys'17* (June 2017), 193–198. https://doi.org/10.1145/3083187.3083210

[24] Matt Yu, Haricharan Lakshman, and Bernd Girod. 2015. A Framework to Evaluate Omnidirectional Video Coding Schemes. In *2015 IEEE International Symposium on Mixed and Augmented Reality*. 31–36.