

Influence of Emotions on Eye Behavior in Omnidirectional Content

Wei Tang^{1,2}, Shiyi Wu^{1,2}, Toinon Vigier¹, Matthieu Perreira Da Silva¹

¹LS2N UMR CNRS 6004, Université de Nantes, France, ²South China University of Technology, China

Abstract—The recent development of Virtual Reality (VR) technologies and interactive visual content faces some important challenges in multimedia processing and Quality of Experience (QoE). Considering omnidirectional, also called 360-degree, content, one major research topic is the development of reliable visual attention models. In this paper, we present a new dataset to study the influence of emotions on eye behavior in omnidirectional content. This dataset is based on an eye-tracking experiment where 19 observers have assessed emotional valence and arousal dimensions in 360-degree images. Several analyses are then conducted to compare fixation and saccade features, inter-observer saliency congruency and spatial oculomotor biases in positive, neutral and negative content. Results show a significant impact of negative images on visual attention, with more visual agitation and avoidance behavior from larger, longer and faster saccades. However, no obvious difference of eye behavior is found between positive and neutral stimuli on this dataset.

Index Terms—eye behavior, emotions, omnidirectional images, eye features, saliency

I. INTRODUCTION AND RELATED WORKS

The recent development of Virtual Reality (VR) technologies and interactive visual content faces some important challenges in multimedia processing and Quality of Experience (QoE). Considering omnidirectional, also called 360-degree content, one major research topic is the development of reliable visual attention models for improving processing, coding delivering and rendering techniques [1].

In this context, Rai et al [2] published the first, to our knowledge, 360-degree image dataset with eye data in order to understand how users watch and explore this new type of immersive content, as well as to benchmark visual attention models. This dataset is accompanied by a toolbox and an evaluation framework to analyze eye features and compare visual saliency maps in 360-degree context [1]. Recently, it was extended to omnidirectional video content [3]. These publications point out the importance of new datasets to study eye movements in immersive and interactive content, as well as the need of new tools to analyze eye data combined with head movement in omnidirectional content visualized in head mounted display (HMD).

Emotion is an important factor of user QoE. Several studies have investigated and confirmed that visual attention and eye movements are altered in the presence of emotional stimuli. Lemos et al. [4] thus, introduced a method to recognize emotions using eye data considering pupil size, blink characteristics, and gaze features. They noted that combining some

features could improve emotion recognition efficiency, but lacks validated data and methods. It has been also shown that human attention usually takes precedence over emotional content rather than neutral content [5]–[7]. Indeed, Nummenmaa et al. [8] recorded gaze data while the visualization of 128 pictures from International Affective Picture System (IAPS) and found that the probability of fixating an emotional image for the first time and the frequency of subsequent gaze are greater than for neutral image. More recently, Wiebe et al. [7] showed an impact of emotion conditions on attention duration. Their results, obtained on 80 different pictures, indicated that participants looked longer at depression-related and happy images than neutral ones, and even longer at anxiety-related pictures.

Bailenson et al. were the first, to our knowledge, to investigate the relationship between emotions and visual attention, through observer's head movements, in omnidirectional content [9]. In their study, they recorded head data during the visualization of several VR clips, which were emotionally scored in valence and arousal with the Self Assessment Manikin (SAM) scale. The results showed a significant relationship between the amount of head yaw and valence score revealing that people who had more side-to-side movements had more positive emotions. A significant correlation between head pitch and arousal was also found. However, the duration of the clips was quite long with strong variations in content type and semantic value, reducing the precision of the emotional scoring and the head features analysis for each content.

Considering all these related works, we think that the study of the influence of emotions on visual behavior in omnidirectional images would be very fruitful for QoE, image processing and affective computing domains. Therefore, we propose to construct and analyze the first, to our knowledge, eye tracking dataset in emotional omnidirectional content.

The rest of the paper is organized as follows. Section II describes the details of the material experiment. Section III is additional description of the dataset mainly about emotional scores and eye features. Section IV portays and analyses the link between eye data and emotions.

II. EXPERIMENT

A. Stimuli

High quality 360-degree images were required for this experiment. At first, neutral and positive images were extracted from an existing 360 saliency dataset [2]. However, this dataset does not contain any negative content. Therefore, negative

images were extracted from videos found on the Internet. The resolutions of the equi-rectangular images range from 2000×1000 pixels to 12000×6000 pixels. Some of them are shown in Fig. 1. In a first time, we selected first 45 images and assessed them in a pre-test with SAM scale. Finally, 36 omnidirectional images were selected for the experiment.

B. Experimental setup

The experiment was performed with a HTC Vive headset equipped with a SMI eye tracker that captures the position of each eye at 75 Hz. The helmet is capable of capturing head rotation data on three axes to deliver the appropriate portion of the scene to the viewer. The original eye movements data recorded by SMI eye tracker include left and right positions (x, y), left and right gaze base points (x, y, z), left and right and average gaze directions (x, y, z).

C. Subjective methodology

The experiment was conducted in a quiet room. Subjects sat in a rolling chair so they could have the freedom to move and rotate 360 degrees. After receiving instructions, subjects performed a pre-test that would help them become familiar with the VR headset, UI menu and controller. The complete test included 36 images in a random order. After each image the observer had to complete the SAM questionnaire for arousal and valence. A full gray image was rendered for 5 seconds before each image played to avoid any bias due to the previous image. Each test person lasted about 20 to 25 minutes.

D. Participants

Twenty-four participants were recruited in the experiment, including 14 males and 10 females aged between 20 and 63. All participants were taken to the screening room to test their vision first. The dominant eye was also determined. Before the test, each participant was asked to sign an experimental consent. After the experiment, each subject was compensated.

III. DESCRIPTION OF THE DATASET

A. Emotional scores

5 outliers were found because of correlation < 0.5 with all the other users on SAM scale, which could be due to a misunderstanding of the emotional questionnaire. Therefore data of nineteen test persons were used finally. In order to see the quality of the database and get the basic emotional information of each image in the database, the mean scores and standard errors of 19 subjects for each image were calculated and shown in Fig. 2. It is found that from an emotional point of view, the emotional trend of the database is reasonable. For negative images, it is always accompanied by a relatively high degree of arousal, and neutral images always have relative low arousal, which makes the whole picture looks like a "V" shape and this trend is similar to the IAPS database [10]. In this work, we aimed to analyse the relationship between different emotional omnidirectional images with eye gaze. Due to few images having high emotional variance we clustered 5 scenes

as positive condition, 5 as neutral and 5 as negative according to valence to better represent the positive, neutral and negative emotions. These 15 images used for the following analysis section are shown in Figure 1.

B. Eye data

1) *Eye features*: Eye features related to fixations and saccades have been widely used in the study of disease diagnosis [11], emotion recognition and psychological processes, such as reading [12] and visual search [13]. Due to the problem of boundary discontinuity in equirectangle projection, we chose to analyze the gaze data of fixations and saccades in the sphere map space [2]. Therefore, the orthodromic distance, instead of the Euclidean distance, and a velocity-based algorithm were used to extract fixations and saccades from gaze data. Several descriptive features like fixations and saccades duration or saccades velocity, fixations location and dispersion were then computed.

2) *Saccadic angle*: Simola et al. [14] showed that emotional content changed the angular behavior of eye movements by measuring the saccadic angle. Here, two types of saccadic angle were computed: absolute angle from the horizontal axis and relative angle between two consecutive saccades.

3) *Saliency maps*: Saliency maps were generated by head and eye information. Taking into account the accuracy of the eye tracker and the foveal sensation, saliency maps are computed by convolving each fixation with a Gaussian distribution with a sigma of 2° [3].

The dataset is publicly available and can be downloaded at ftp://ftp.ivc.polytech.univ-nantes.fr/EMOTION_EYE_360.

IV. ANALYSIS AND DISCUSSION

From emotional and gaze data described in the preceding section, we conducted different analysis to explore the influence of emotions on visual behaviour. Firstly, statistical analysis was used to compare the differences of extracted features and the possible connection between eye characteristics and emotions. Then, we observed the influence of emotional content on inter-observer congruence and on spatial oculomotor biases. Finally, we compared our results to an existing dataset to confirm that the obtained results are not due to other content features.

A. Statistical analysis of eye features

A Kruskal-Wallis one-way analysis was used to research the significant effects of eye features between different types of 360-degree images. Results are shown in Table I and II.

1) *Fixations*: Researches have found that visual attention is more focused on emotional pictures and emotional objects than neutral [8]. Our results show a significant impact of perceived emotion on number of fixation points ($\chi^2 = 7.87$, $p = 0.0195$) and fixation duration ($\chi^2 = 18.06$, $p = 0.001$), with more and longer fixation points in positive (number: $M = 70.18$, $SE = 1.33$; mean duration: $M = 273.40$, $SE = 7.44$) and neutral images (number: $M = 67.53$, $SE = 1.30$; mean duration: $M = 280.77$, $SE = 7.45$) than negative ones



Fig. 1. 15 representative omnidirectional images of negative (left column), neutral (second column) and positive (third column) emotions with Valence/Arousal.

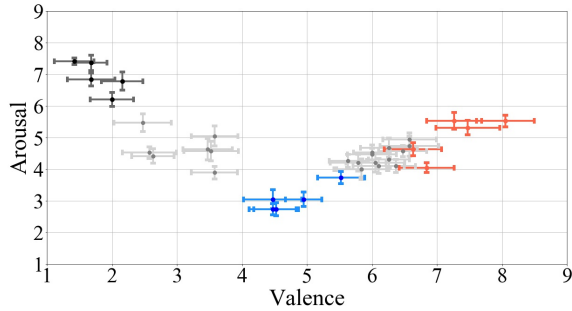


Fig. 2. SAM scores of 36 images. For clustering, we selected the 5 with the lowest valence as Negative (black); the 5 with the highest valence value as Positive (red); the 5 with the lowest arousal as Neutral (blue).

(number: $M = 64.61$, $SE = 1.35$; mean duration: $M = 279.70$, $SE = 7.57$) (see Figure 3). Post hoc comparisons (see Table II) indicate a significant difference only between positive and negative groups. Observers did more fixations and spent more time attentively watching content in pleasant images. It can be explained by the fact that positive images tend to put the observers in a state of contemplation.

2) *Locations*: Fig. 3 shows that location and dispersion on Y-axis are significantly higher in negative images (location Y: $\chi^2(2, 282) = 17.9186$, $p < 0.001$; dispersion Y: $\chi^2(2, 282)$

TABLE I
KRUSKAL-WALLIS TEST RESULTS

Features	χ^2	p-Value
Number of fixations	7.87	0.0195*
Fixation duration SUM	18.06	0.001**
Fixation duration MEAN	0.96	0.6188
Fixation duration STD	8.5524	0.0139*
Fixation duration MAX	5.8295	0.0542
Location X	1.6152	0.4459
Location Y	17.9186	0.0001**
Dispersion X	0.6216	0.7329
Dispersion Y	46.9343	<0.0001***
Saccade duration MEAN	31.7498	<0.0001***
Saccade duration STD	36.9276	<0.0001***
Saccade duration SUM	21.88343	<0.0001***
Saccade duration MAX	31.1307	<0.0001***
Saccade amplitude MEAN	80.3720	<0.0001***
Saccade amplitude SUM	36.3234	<0.0001***
Saccade amplitude STD	76.7444	<0.0001***
Saccade amplitude MAX	42.3475	<0.0001***
Saccade velocity MEAN	21.4895	<0.0001***
Saccade velocity MAX	5.3161	0.0701
Absolute angle MEAN	0.0290	0.9856
Absolute angle MAX	4.5331	0.1037
Relative angle MEAN	1.6920	0.4291
Relative angle MAX	1.5293	0.4655

$= 46.9343$, $p < 0.0001$). Post hoc comparisons indicate that negative condition (location Y: $M = 0.56$, $SE = 0.0074$; dispersion Y: $M = 0.11$, $SE = 0.0031$) is significantly different

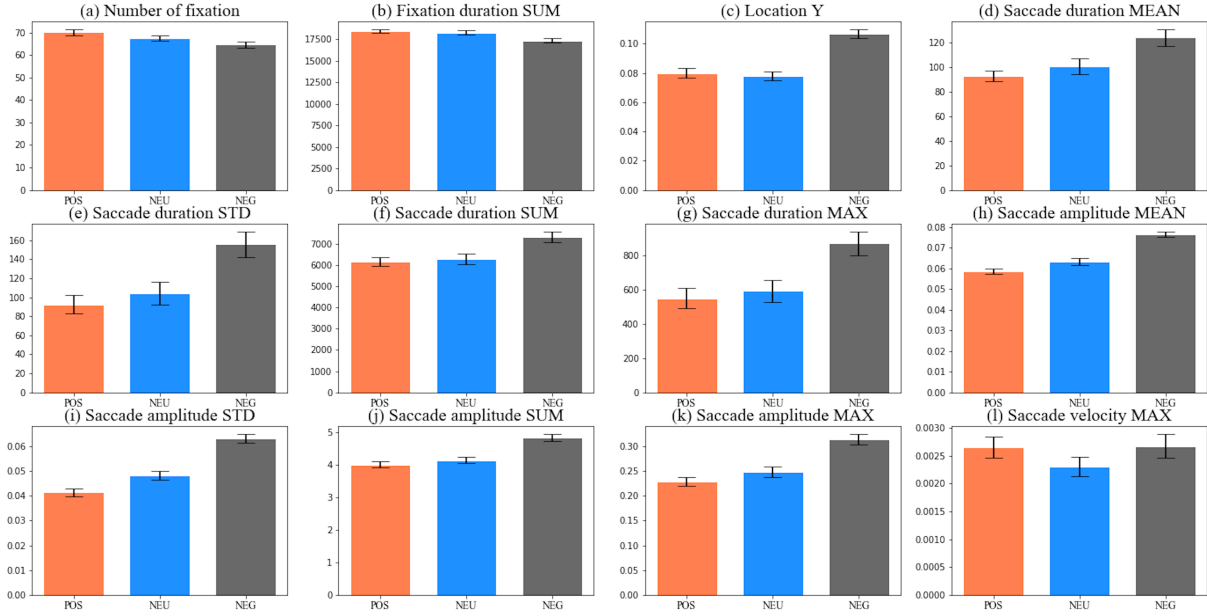


Fig. 3. Mean and standard errors of different eye features.

TABLE II
POST HOC RESULTS

Features	p(Pos-Neu)	p(Pos-Neg)	p(Neu-Neg)
Number of fixations	0.4433	0.0142*	0.2538
Fixation duration SUM	0.9895	0.0005**	0.0009**
Fixation duration STD	0.0254*	0.0369*	0.9897
Location Y	0.9971	0.0008**	0.0006**
Dispersion Y	0.8377	<0.0001***	<0.0001***
Saccade duration MEAN	0.5402	<0.0001***	<0.0001***
Saccade duration STD	0.5199	<0.0001***	<0.0001***
Saccade duration SUM	0.9981	<0.0001***	<0.0001***
Saccade duration MAX	0.8331	<0.0001***	<0.0001***
Saccade amplitude MEAN	0.0959	<0.0001***	<0.0001***
Saccade amplitude STD	0.0126	<0.0001***	<0.0001***
Saccade amplitude SUM	0.5992	<0.0001***	<0.0001***
Saccade amplitude MAX	0.5262	<0.0001***	<0.0001***
Saccade velocity MEAN	0.8662	<0.0001***	0.0005**

from positive (location Y: $M = 0.51$, $SE = 0.0038$, $p < 0.0008$; dispersion Y: $M = 0.08$, $SE = 0.0033$, $p < 0.0001$;) and neutral (location Y: $M = 0.51$, $SE = 0.0041$, $p < 0.0006$; dispersion Y: $M = 0.077$, $SE = 0.003$, $p < 0.0001$;) conditions. A higher value of location Y could suggest a more back slanted position during the visualization of negative content due to less involvement in the content. Concerning dispersion, results indicate a larger exploration in negative content, which could be explained by the avoidance of negative part of the images. This result is in contradiction with a previous study which showed that inducing positive stimuli broadens the visual attention, while negative emotional stimuli narrows attention in traditional 2D visual content [15].

3) *Saccades*: The features associated with saccades are significantly different ($p < 0.0001$), except for maximum saccade velocity ($p = 0.0701$) which is mainly due to physiological

limitations. From the result of saccades duration in Fig. 3, it can be seen that the average ($M = 124.19$, $SE = 6.91$), variance ($M = 155.55$, $SE = 13.25$), maximum ($M = 871.59$, $SE = 68.74$), and total ($M = 7325.76$, $SE = 7325.76$) values of the duration of negative images are greater than the positive (average: $M = 92.92$, $SE = 4.42$; variance: $M = 92.60$, $SE = 10.05$; maximum: $M = 551.58$, $SE = 57.96$; total: $M = 6160.71$, $SE = 216.48$) and neutral ones (average: $M = 100.71$, $SE = 6.24$; variance: $M = 104.39$, $SE = 12.17$; maximum: $M = 64.30$, $SE = 64.30$; total: $M = 6306.03$, $SE = 250.44$). Post hoc comparisons show that each group of features has a significant difference between negative and neutral, negative and positive, while the positive and neutral differences are not significant (see Table II). The results for saccade amplitude are similar to duration: the negative images have a higher amplitude and are significantly different from the positive and neutral values. These results are consistent with a previous study. As mentioned in [14], saccades are significantly longer during the visualization of strongly negative or threat-based stimuli compared to pleasant and neutral images. As previously shown by [16], we also found that the speed of saccades depends on the emotion of the visualized content. The average saccadic velocity for negative content was significantly higher than for positive and neutral images, and the result is significant (Pos-Neg: $p < 0.0001$, Neu-Neg: $p < 0.001$). Statistical results show longer, larger and faster saccades in negative content which could reflect an agitated mental state.

Statistical analysis of saccade and fixation features show that visual behavior is affected by negative omnidirectional content with less numerous, more dispersed and shorter fixations along with longer and faster saccades. Although these results do not really confirm the previous results obtained in still 2D

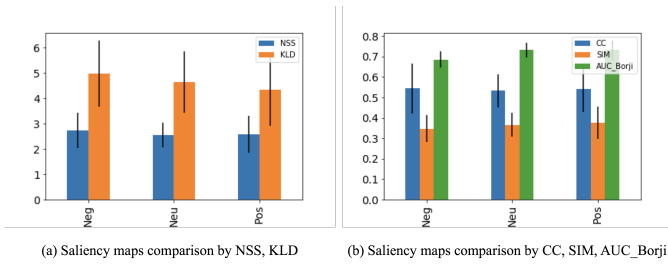


Fig. 4. Mean and standard deviations of saliency maps comparison.

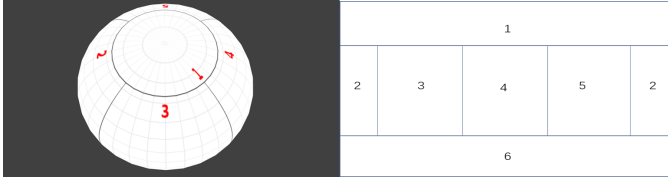


Fig. 5. The way to divide one 360-degree image for analysis.

images, they look consistent altogether and be interpreted as a more agitated and less attentive visual behavior during the visualization of omnidirectional images inducing negative feelings. We may assume that negative contents in VR HMD are quite impressive, thus they can induce discomfort and avoidance behaviour.

B. Analysis of saliency maps

We may also assume that this more agitated behavior in negative content can lead to a reduction of the visual attention congruence between all the observers. In order to confirm or infirm this hypothesis, we compared the saliency map of each observer to each image with the average saliency map of this specific image. Two equirectangular saliency maps cannot be directly compared as in a typical 2D context because the poles of the sphere are over-represented at the top and bottom of the equirectangular image. In other words, there is too much data at the pole compared to the data at the equator. Thus the method described in [2] was used to compute the similarity between saliency maps thanks to the following metrics: linear normalized scanpath saliency (NSS), similarity metric (SIM), correlation coefficient (CC), area under the ROC curve (AUC-Borji and Kullback-Leibler Divergence (KLD). These indicators can measure different aspects including distribution and location [17] between observers.

Results in Fig 4 show the mean and standard deviations of these metrics for negative, neutral and positive content. It should be pointed that the results is ranked by negative, neutral and positive every 5 images. Even if metrics show a tendency to lower consistency between observers for negative images, differences are slight and we cannot validate our hypothesis.

C. Impact of emotions on spatial oculomotor biases

In order to extend the results on fixations location, we further explore spatial biases under different emotional stimuli.

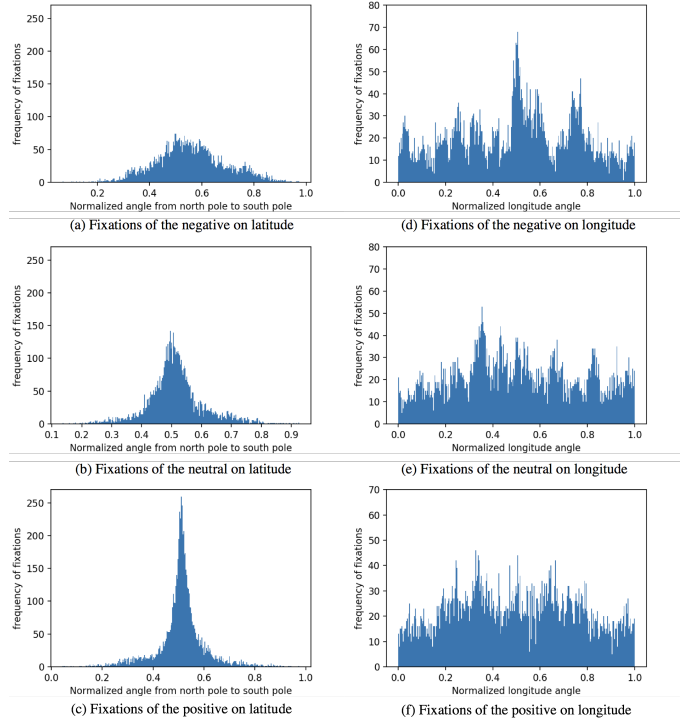


Fig. 6. Longitude and altitude spatial biases comparison.

First we compared horizontal and central biases in all emotional condition (see Fig. 6), and found that positive images have the strongest center bias in latitude which is coherent with a shorter Y-dispersion of fixations in positive images.

Then, we deeply investigated spatial biases by dividing the sphere in 6 parts as shown in Fig. 5 and computed the saccade absolute angle and amplitude distribution in each area. Absolute angle represents the angle between saccadic vector and the equator. Density probability polar distributions of saccade absolute angle and amplitude are represented in Fig. 7. Except in up-area (area 1) where we can find more upward vertical in negative content, we could not find differences in oculomotor biases according to emotional content.

D. Comparison with salient360 dataset

The statistical analysis of gaze data features shows significant differences between negative and more pleasant images. However, our dataset is quite small and all negative images are indoor whereas positive ones are outdoor. Therefore, in order to ensure that results are not due to the type of content, we performed the same analysis on a publicly available eye-tracking dataset on omnidirectional images¹. From this dataset, we selected two subsets containing five outdoor images (P11, P38, P54, P58, P88) and five indoor images (P10, P22, P28, P82, P97). Then, as we did for emotional content, we applied statistical test to investigate significant effect of type of content on gaze data features. Because the provided data do not contain

¹<https://salient360.ls2n.fr/datasets/training-dataset/>

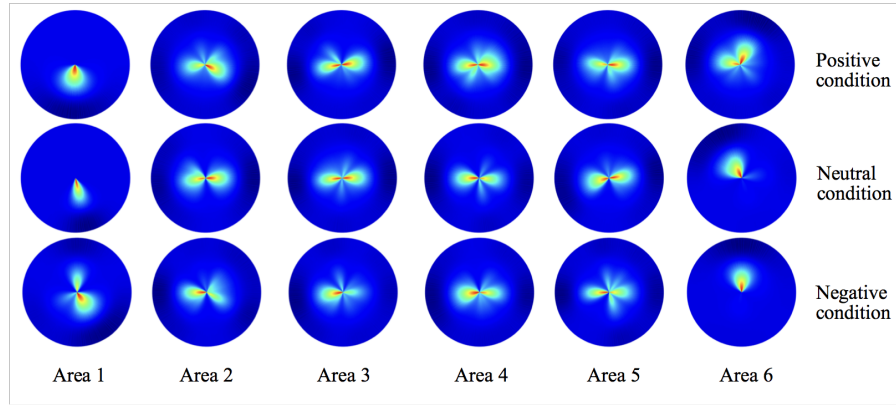


Fig. 7. Density probability distributions of absolute angle and saccade amplitude (deg.) represented as polar plots.

duration information for fixations and saccades, only number of fixations, fixation location, saccade amplitude and absolute angle were compared.

Results in III show that all the features, except for fixation location on Y-axis, are not significantly different, implying that most of the results in previous analyses appear to be due to emotions. However, they should be confirmed in a further study on a larger dataset.

TABLE III
STATISTICAL RESULTS FOR INDOOR/OUTDOOR SUBSETS COMPARISON

Features	P-Value
Number of fixations	0.4489
Location X	0.3166
Location Y	<0.0001***
Dispersion X	0.5779
Dispersion Y	0.9387
Saccade amplitude(mean)	0.5241
Saccade amplitude(std)	0.6610
Saccade amplitude(max)	0.5488

V. CONCLUSION AND FUTURE WORKS

In this paper, we have presented a new eye tracking with valence / arousal scores dataset on 36 omnidirectional content visualized by 19 observers in a VR HMD device. This dataset has been used to study the influence of emotions on eye behavior in immersive content. Several statistical and exploratory analyses were then conducted. Results show a significant impact of negative emotions on fixation and saccade features, which can be interpreted as a more agitated and less attentive visual behavior due to discomfort and avoidance behaviour. However, even if these promising results have been validated by comparison with another dataset, they should be confirmed in new omnidirectional datasets with more diverse content.

REFERENCES

- [1] J. Gutiérrez, E. J. David, A. Coutrot, M. P. Da Silva, and P. Le Callet, "Introducing un salient360! benchmark: A platform for evaluating visual attention models for 360 contents," in *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2018, pp. 1–3.
- [2] Y. Rai, J. Gutiérrez, and P. Le Callet, "A dataset of head and eye movements for 360 degree images," in *Proceedings of the 8th ACM on Multimedia Systems Conference*, 2017, pp. 205–210.
- [3] E. J. David, J. Gutiérrez, A. Coutrot, M. P. Da Silva, and P. L. Callet, "A dataset of head and eye movements for 360 videos," in *Proceedings of the 9th ACM Multimedia Systems Conference*, 2018, pp. 432–437.
- [4] J. De Lemos, G. R. Sadeghnia, Í. Ólafsdóttir, and O. Jensen, "Measuring emotions using eye tracking," in *Proceedings of measuring behavior*, vol. 226, 2008, pp. 225–226.
- [5] T. Brosch, G. Pourtois, and D. Sander, "The perception and categorisation of emotional stimuli: A review," *Cognition and emotion*, vol. 24, no. 3, pp. 377–400, 2010.
- [6] P. Vuilleumier, "How brains beware: neural mechanisms of emotional attention," *Trends in cognitive sciences*, vol. 9, no. 12, pp. 585–594, 2005.
- [7] A. Wiebe, A. Kersting, and T. Suslow, "Deployment of attention to emotional pictures varies as a function of externally-oriented thinking: An eye tracking investigation," *Journal of behavior therapy and experimental psychiatry*, vol. 55, pp. 1–5, 2017.
- [8] L. Nummenmaa, J. Hyönä, and M. G. Calvo, "Eye movement assessment of selective attentional capture by emotional pictures," *Emotion*, vol. 6, no. 2, p. 257, 2006.
- [9] B. J. Li, J. N. Bailenson, A. Pines, W. J. Greenleaf, and L. M. Williams, "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*, vol. 8, p. 2116, 2017.
- [10] P. Lang, M. Bradley, and B. Cuthbert, "International affective picture system (iaps): affective ratings of pictures and instruction manual. university of florida, gainesville," Tech Rep A-8, Tech. Rep., 2008.
- [11] H. M. Livengood and N. A. Baker, "The role of occupational therapy in vision rehabilitation of individuals with glaucoma," *Disability and rehabilitation*, vol. 37, no. 13, pp. 1202–1208, 2015.
- [12] O. Kardan, M. G. Berman, G. Yourganov, J. Schmidt, and J. M. Henderson, "Classifying mental states from eye movements during scene viewing," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 41, no. 6, p. 1502, 2015.
- [13] J. M. Henderson, S. V. Shinkareva, J. Wang, S. G. Luke, and J. Olejarczyk, "Predicting cognitive state from eye movements," *PloS one*, vol. 8, no. 5, 2013.
- [14] J. Simola, K. Le Fevre, J. Torniaainen, and T. Baccino, "Affective processing in natural scene viewing: Valence and arousal interactions in eye-fixation-related potentials," *NeuroImage*, vol. 106, pp. 21–33, 2015.
- [15] H. Van Steenbergen, G. P. Band, and B. Hommel, "Threat but not arousal narrows attention: evidence from pupil dilation and saccade control," *Frontiers in psychology*, vol. 2, p. 281, 2011.
- [16] J. M. Susskind, D. H. Lee, A. Cusi, R. Feiman, W. Grabski, and A. K. Anderson, "Expressing fear enhances sensory acquisition," *Nature neuroscience*, vol. 11, no. 7, p. 843, 2008.
- [17] O. Le Meur and T. Baccino, "Methods for comparing scanpaths and saliency maps: strengths and weaknesses," *Behavior research methods*, vol. 45, no. 1, pp. 251–266, 2013.