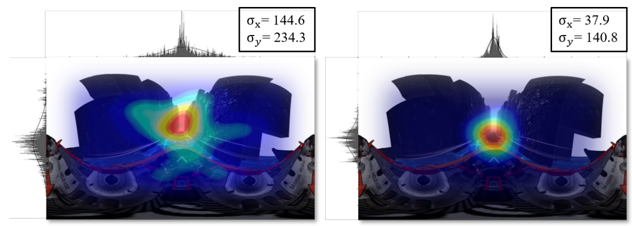
A Study on Virtual Reality Sickness and Visual Attention

Jeonghaeng Lee*, Woojae Kim*, Jinwoo Kim*, and Sanghoon Lee*†
* Department of Electrical and Electronic Engineering, Yonsei University, Seoul, South Korea
† Department of Radiology, College of Medicine, Yonsei University, Seoul, South Korea
E-mail: {leedoright, wooyoa, jw09191, slee}@yonsei.ac.kr, Tel: +82-2-2123-2767

***Abstract*—It is a well-known fact that virtual reality (VR) sickness is an obstacle to an immersive VR experience, however, an objective analysis of the physiological responses for VR sickness has been insufficient. In this study, our analysis uncovers how the users' visual attention varies with the level of VR sickness and how the level of VR sickness influences the center-bias tendency. Toward this, we first conduct a large-scale eye-tracking experiment of 21 inexperienced users while they experience VR sickness-oriented database VR-SP [15]. Then, we quantify the tendency of visual behavior according to VR sickness. To do this, we newly define a visual entropy measurement of VR visual attention. The experimental results clearly suggest that the center-bias effect becomes stronger as the degree of VR sickness increases. In other words, this implies that the users' *explorativeness* in VR content may be restricted by the VR sickness and this leads to the restraint of the immersive experience. For a more clear demonstration, we also show the visual entropy can be used to predict VR sickness with an accuracy of 80% on the VR-SP database.**



(a) Sickness: 0.04 (b) Sickness: 0.65

Fig. 1: Examples of eye-tracker results according to their VR sickness levels. (a) is low sickness example and (b) is a high sickness example of the "AirFighter" scene in the VR-SP database.

I. Introduction

The virtual reality (VR) industry as a whole is advancing at a fast pace and also has the potential to have a notable impact on our society. By extending the freedom of movement of a rigid body in three-dimensional space, the user expects an increasingly high quality of experience (QoE) from VR services. Therefore, an unexpected experience through a head-mounted display (HMD) may induce a deleterious physiological side-effect, *i.e.*, *VR sickness*, which prevents the promotion of related industrial development [19].

In early studies, the physiological mechanism of motion sickness which is the higher concept to explain VR sickness has been analyzed from a neuro-science field [13]. Reason *et al.* postulated that motion sickness was caused by internal- and intra-sensory conflicts between visual, vestibular, and somatosensory [18]. Bles *et al.* clarified this theory by employing a vector sum of gravity and inertial acceleration by modeling the sensed vertical [1]. Bos *et al.* expanded the above theory to a dynamic model with a mathematical description [2]. However, the theories are only shown analytically, and VR-oriented experimental analysis has been still insufficient yet.

More recently, deep-learning techniques have been successfully applied to estimate human perceptual QoE [11], [14], [16]. Inspired by this, Padmanaban *et al.* [17] proposed a first machine-learning approach to predict VR sickness. Since a motion component is assertive to VR sickness, several motion-estimation methods have been used to predict the subjective VR sickness. Kim *et al.* [8] developed a sickness prediction model to generate an exceptional motion map that is defined as a difference between image input and reconstructed output. Kim *et al.* [12] designed a novel architecture that imitates and learns the neurological mechanism of motion sickness.

As the previous studies emphasized [8], [12], [17], it is a well-known fact that visual information occupies superiority among various sensory information in a VR environment. The user tends to comprehend the graphic domain surrounding the user itself through the visual information and predicts the expected movement for external environment changes. In this respect, the visual attention on each VR content should be regarded as one of the most objective ways to reflect VR sickness. Fig. 1 shows the representative tendency of visual attention according to VR sickness. Figs. 1 (a) and (b) are the two different VR contents overlaid with their visual attention map, *i.e.*, (b) has higher VR sickness score than (a). Through the figure, we found that the users' visual attention tends to become strongly center-biased as the VR sickness intensifies. Conversely, when VR sickness is slight, users' *explorativeness* is much larger. This inspires us that the degree of VR sickness can be predicted by quantifying the distribution of visual attention.

In this study, we perform a comprehensive analysis of the users' visual attention in the VR environment according to the degree of VR sickness. To do this, we conduct eye-tracking experiments on a large-scale VR sickness-dedicated database [15]. Moreover, to quantify the visual attention, we newly devise an information-theory-based entropy measurement for the

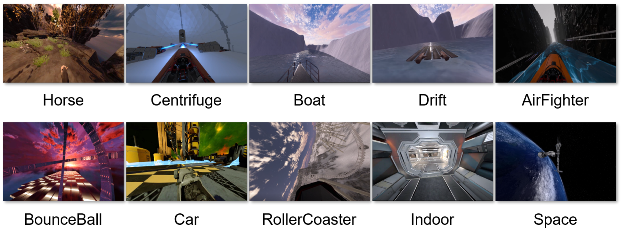


Fig. 2: Exemplar reference scenes in the VR-SP database [15].

VR environment. Furthermore, we show the proposed entropy of visual attention highly correlates with users' subjective VR sickness scores.

II. VR EYE-TRACKING EXPERIMENT

A. VR content stimuli

For the eye-tracking experiment reported in this paper, we use the VR-SP database that contains 100 individual VR contents with corresponding VR sickness subjective scores [15]. The VR-SP database consists of 10 reference VR scenes while each reference includes 10 sub-scenes having various degrees of VR sickness. All reference scenes are designed by artists in the Unity platform. Note that various degrees of VR sickness in each referenced scene are designed by combinations of two texture types, two motion types, and four levels of velocity. Fig. 2 shows examples of 10 reference scenes in the VR-SP database. We use the 360-degree video version of VR sequences in the VR-SP database, and those VR contents have a resolution of 2880×1600 and a duration of 10 seconds.

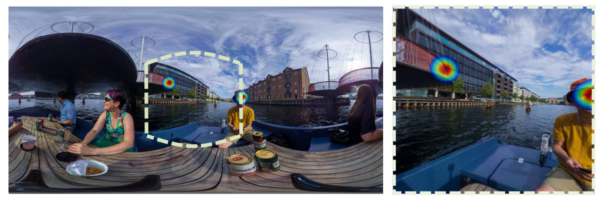
B. Eye-tracking procedure

In the experiment, a total of 21 inexperienced subjects (satisfying the subject criteria recommended in [3]) was participated (17 male, 4 female, ages ranging from 22 to 36). They were seated in a swivel chair while wearing a HMD device. The overall eye-tracking protocol included four recording sessions, each containing 25 randomly shuffled VR content stimuli from the database. The Unity game engine was utilized to display all protocols and record head orientation while the eye tracker collected eye movement. For each subject, the eye-tracker was calibrated using six points calibration at the beginning of the protocol. To consistently initialize the fixation point of each stimulus, subjects were instructed to watch a fixation cross at the center of a gray screen for 5 seconds. After each session (25 VR contents), rest periods of 5 min duration were inserted, to minimize any accumulated VR sickness [22]. We summarize major information about the test environment in Table I.

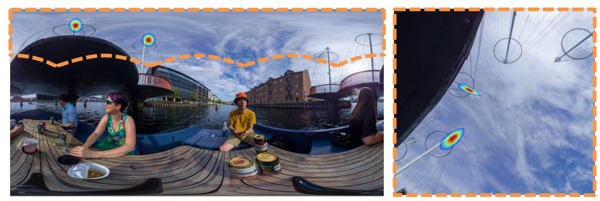
In each session, each VR content stimulus was displayed for 10 seconds while the eye movement was recorded by the built-in eye-tracker of HTC-VIVE PRO EYE. Here, the subjects were asked to *freely look around the VR content* without any unnecessary restrictions. After recording all subjects' eye movements, 3D fixation angles were obtained by combining

Simulation platform	Unity
Color depth	24-bits/pixel color frames
Video coder	MPEG-4
Subjects	21 inexperienced subjects
Frame resolution	2880×1600
Duration	10 seconds (60 fps)
# of VR content	100
Viewing environment	HTC-VIVE PRO EYE
# of fixations per content	12600

TABLE I: Configurations and conditions of the eye-tracking experiment.



(a) Front region



(b) Top region

Fig. 3: Examples of latitudinal distortions in equirectangular image. (a) shows the front region distortions in users field of view and (b) show the top region example.

the head orientation vectors (pitch, yaw, and roll) and gaze direction vectors *w.r.t.* each VR content.

III. VISUAL ENTROPY MEASUREMENT

A. Spherical heatmap generation

To quantify the visual attention in the VR environment, we model a statistical distribution of the fixation points. To this end, we use a Gaussian mixture model (GMM) on the spherical coordinates. The distribution of fixation angles $(\theta_i^t \cdot \phi_i^t)$ of t^{th} frame can be defined in spherical domain as

$$S^t(\theta,\phi) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2\pi\sigma^2}\exp\left(-\frac{(\theta-\theta_i^t)^2+(\phi-\phi_i^t)^2}{2\sigma^2}\right), \quad (1)$$

where N is the number of subjects, and θ_i $(0 \leq \theta_i \leq 2\pi)$ and ϕ_i $(0 \leq \phi_i \leq \pi)$ are 3D fixation angles of i^{th} subject.

To derive a 360 degree based fixation distribution model, we first project the spherical fixation into the equirectangular domain. Here, we denote the projected heatmap as $f^t(x,y)$ where (x,y) is the pixel position in Cartesian coordinates and t is the frame index. Finally, for every 10 seconds of VR

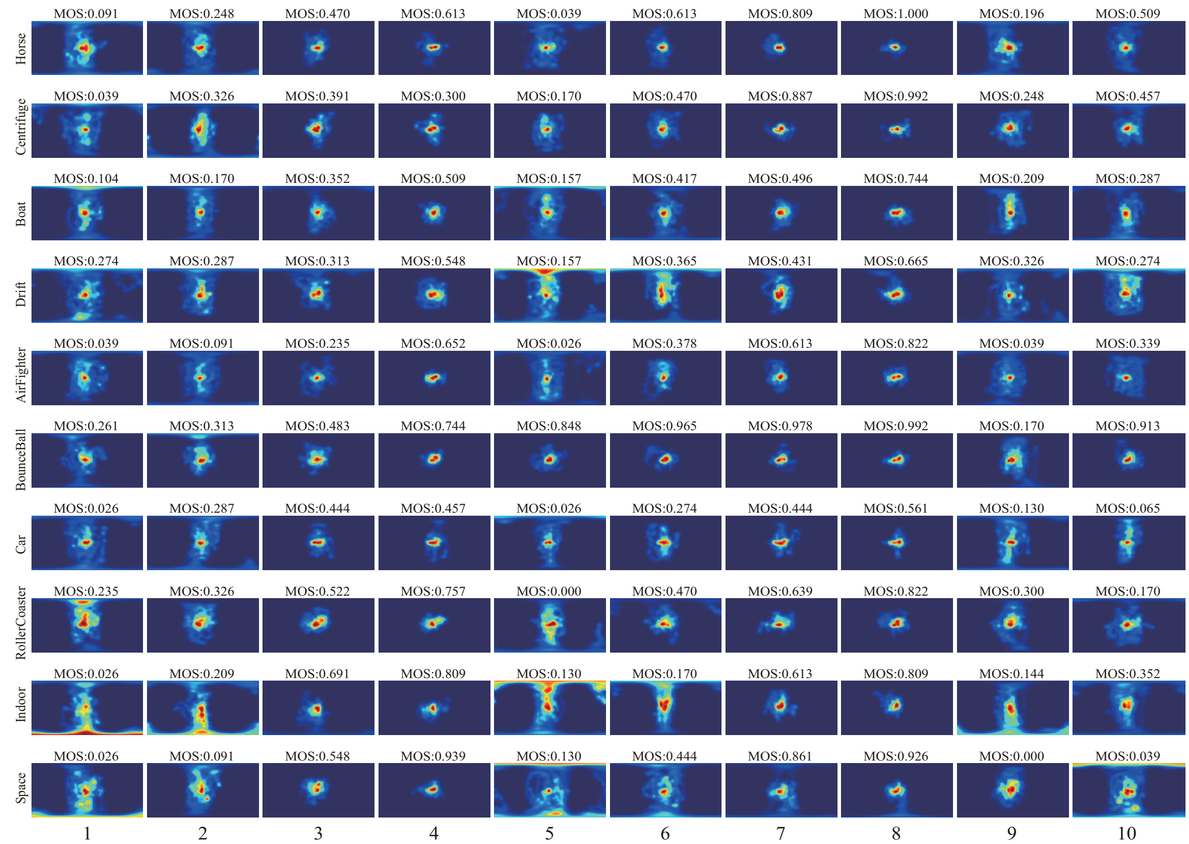


Fig. 4: Equirectangular heatmap of eye-tracking experimental results and the corresponding sickness mean opinion score (MOS).

content, a total of 600 heatmaps is obtained from the eye-tracking results. Fig. 4 reports the equirectangular heatmaps accumulated over 600 frames and corresponding VR sickness scores for all samples in the database. As aforementioned, as the VR sickness score increases, the center-bias tendency more clearly stands out. However, for the lower VR sickness samples, it may be seen that users' fixation distributes wider.

B. Cube-based entropy measurement

It is also important to quantitatively analyze how the fixation distribution varies when VR sickness changes. To do this, Kim *et al.* [6] introduced a multi-scale entropy measurement method for the 2D space. However, in the case of the VR domain, it is necessary to consider the latitudinal distortions. Fig. 3 shows examples of latitudinal distortions in the front and top regions of the equirectangular image. To address this, we newly define a multi-scale cube-based entropy measurement that is well-suited in a VR environment by converting the equirectangular domain into the cubic representation.

Fig. 5 shows a schematic pipeline of cube-based multi-scale distribution modeling for the entropy measurement. We first convert a given equirectangular heatmap $f^t(x, y)$ into the six faces (back, down, front, left, right, and top) by projection method [4]. Then a set of faces are split into multiple scales s. Followed by the threshold of relative object size (ROS) defined

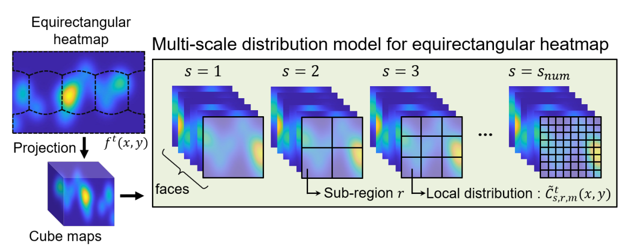


Fig. 5: Multi-scale distribution model for equirectangular heatmap.

in [6], [21], we set the number of scales as 20 (ROS $>$ 5%). We denote the local heatmap distribution for each frame as $C^t_{(s,r,m)}(x, y)$ where r and m are the indices of sub-region and face, respectively. The entropy for each scale, region, and face can be defined by

$$H^t_{(s,r,m)} = -\sum_{x,y} \tilde{C}^t_{s,r,m}(x, y) \log \tilde{C}^t_{s,r,m}(x, y). \quad (2)$$

Finally, the overall entropy of t^{th} frame is expressed as

$$H^t = \frac{1}{s_{num}} \sum_{s,r,m} H^t_{s,r,m}, \quad (3)$$

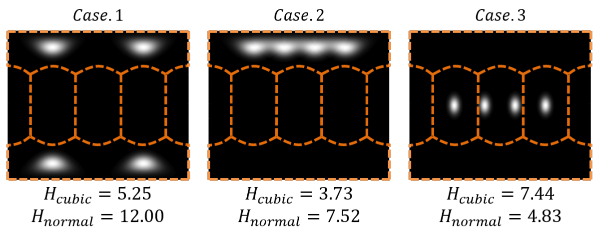


Fig. 6: Comparisons of the normal entropy and proposed cube-based entropy for various fixation maps in equirectangular domain.

where s_{num} is the number of the scale.

Fig. 6 shows a comparison of the proposed cubic entropy and the normal entropy calculation [6] using four fixation heatmaps. In the figure, the dotted lines represent the boundaries of top, bottom, front, back, left, and right, as depicted in the equirectangular to cube-map projection shown in Fig.5. In the figure, the white circles of case 1 are widely distributed over the equirectangular image. However, the actual fixation points to be observed through the HMD are close together because of latitudinal distortion. Therefore, when the proposed cubic entropy is calculated, a lower entropy may be obtained. However, the normal entropy calculation shows the opposite tendency because this only considers 2D spatial signals without considering the topology of the equirectangular domain. Similarly, in case 2, since the fixation points are narrowly distributed only in the top region of the VR environment, the cubic entropy results in a low value. On the other hand, case 3 shows a narrow distribution. However, in the VR environment, the fixation points occupy multiple regions from left, right, and front. In this regard, it is noteworthy that the cubic entropy thoroughly reflects the visual information experienced by the user, resulting in a high entropy value.

IV. Experimental results

A. Temporal analysis

As aforementioned, the visual attention of VR content is highly related to VR sickness. To analyze temporal visual attention, we compare the obtained cubic entropy with the motion feature which is highly correlated with VR sickness. For the motion feature, the average magnitude of Horn-schunck method [5] is used. Fig. 7 shows the temporal distributions of the motion features and the cubic entropy of all scenes in the VR-SP database. As described in Section II-B, the fixation of all subjects is initialized to the center region. Therefore, the entropy at the start of each scene is lowered. When the sequence has low visual movement (*i.e.,* having lower VR sickness) such as #1, #2, #5, #6, #9, and #10, it can be seen that the obtained entropies relatively increased since the subjects seek to explore the VR environment. Conversely, in the rest of the scenes with relatively high visual movement (*i.e.,* having higher VR sickness), the entropies significantly decreased. From these observations, we conclude that visual attention with high VR sickness strongly induces center-biased.

Method	VR-SP	
	PLCC	SROCC
Optical flow-based [5]	0.565	0.575
VRSP [9]	0.683	0.756
Kim *et al.* [10]	0.836	0.833
Kim *et al.* [7]	0.797	0.712
VRSA-NET [8]	**0.843**	0.814
Entropy-based (ours)	0.830	**0.834**

TABLE II: PLCC and SROCC comparisons of VR sickness prediction models on the VR-SP database.

B. Benchmark results

To validate the correlation between the entropy extracted from the visual attention and VR sickness scores, we further compare the performance of the correlation coefficient on the VR-SP database [15]. Here, we employ two standard measures: Pearson's linear correlation coefficient (PLCC) and Spearman's rank-order correlation coefficient (SROCC) which are significant indicators for regression tasks such as quality and discomfort predictions. The total entropy of each scene is obtained by averaging overall frame entropy as

$$H_{total} = \frac{1}{T}\sum_{t=1}^{T} H^t, \tag{4}$$

where T is the total number of frames in a sequence. To fit the scale of the entropy to VR sickness scores, the mean entropy value H_{total} of each sequence is regressed onto the subjective score by using four parameter-based logistic fitting [20]:

$$score = \zeta_2 + \frac{\zeta_1 - \zeta_2}{1 + exp\Big\{-\Big(H_{total} - \frac{\zeta_3}{|\zeta_4|}\Big)\Big\}}, \tag{5}$$

where the model parameters $\zeta_1, \zeta_2, \zeta_3$, and ζ_4 are obtained by the least squared error between the H_{total} and the subjective VR sickness scores.

We benchmark the performance of those five following existing models. For the optical flow model, we use the average magnitude of the following method [5]. For the VRSP [9], Kim *et al.* [7] and Kim *et al.* [8], we implement as the same setup with their original work. In the case of Kim *et al.* [10], since we don't have the brain signal data, only the visual features are used to obtain the prediction score.

Table II shows the performance comparison. As shown in the table, the performance of entropy shows a competitive performance than those of previous content-based VR sickness predictors. This result shows how clearly users' explorativeness is affected by the degree of VR sickness. In other words, as the VR sickness increases, the visual attention is highly center-biased by excessive visual movement, hence it implies that the users' immersive experience also can be limited. These experimental results suggest that not only the features obtained from the content signal but also the physiological response such as the user's eye movement can be used as an important factor in VR sickness assessment.

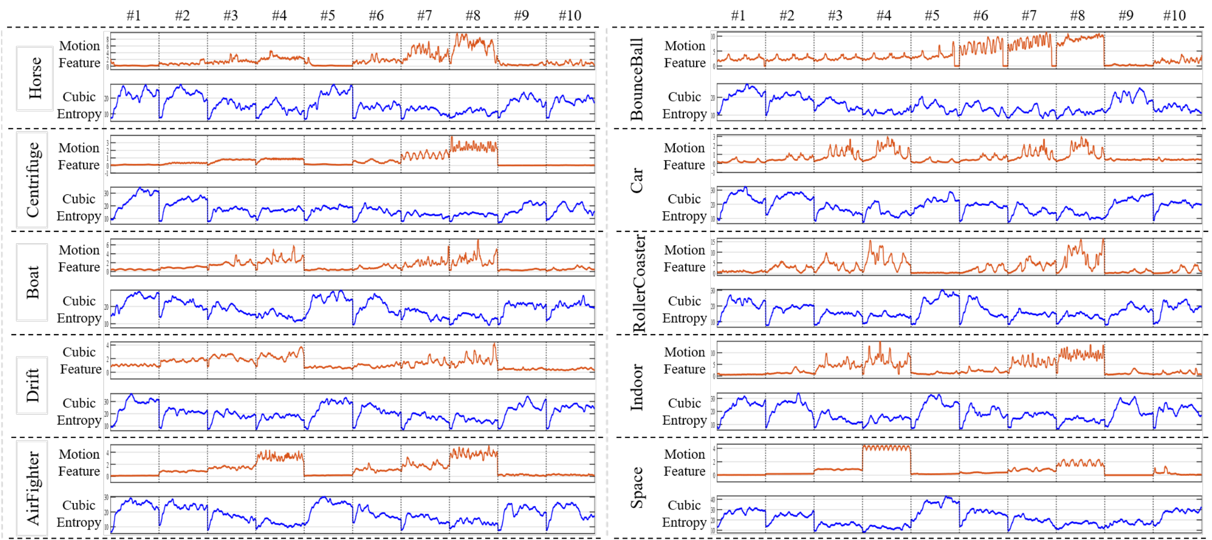


Fig. 7: Temporal distribution of motion feature and cubic entropy in the VR-SP database. #1 to #10 indicate ten individual sub-scenes (duration of 10 sec) of each reference VR scene having diverse degrees of VR sickness.

V. Conclusion

Recently, virtual experience technology has improved remarkably, which raises the importance of understanding VR sickness in commercial HMD products. In this study, by measuring visual entropy in a VR environment, we analyzed how users' visual attention changes according to the degree of VR sickness. From the experimental results, we verified the users' visual attention is highly restricted as the VR sickness increases, *i.e.,* immersive experience in VR content is also limited in higher VR sickness. In other words, the eye-tracking results of users in a VR environment can be used as a significant factor to evaluate subjective or objective VR sickness assessment.

Acknowledgment

This work was supported by Institute of Information communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2021-0-00352, Development of Content Quality Test Method Based on Descriptive Sensible Experience Prediction Model)

References

[1] W. Bles, J. E. Bos, B. De Graaf, E. Groen, and A. H. Wertheim. Motion sickness: only one provocative conflict? *Brain Research Bulletin*, 47(5):481–487, 1998.
[2] J. E. Bos, W. Bles, and E. L. Groen. A theory on visually induced motion sickness. *Displays*, 29(2):47–57, 2008.
[3] R. I.-R. BT.500. Methodology for the subjective assessment of the quality of television pictures. 2002.
[4] N. Greene. Environment mapping and other applications of world projections. *IEEE Comput. Graph. Appl.*, 6(11):21–29, 1986.
[5] B. K. Horn and B. G. Schunck. Determining optical flow. In *Techniques and Applications of Image Understanding*, volume 281, pages 319–331. International Society for Optics and Photonics, 1981.
[6] H. Kim and S. Lee. Transition of visual attention assessment in stereoscopic images with evaluation of subjective visual quality and discomfort. *IEEE Trans. Multimedia*, 17(12):2198–2209, 2015.
[7] H. G. Kim, W. J. Baddar, H.-t. Lim, H. Jeong, and Y. M. Ro. Measurement of exceptional motion in VR video contents for VR sickness assessment using deep convolutional autoencoder. In *Proc. ACM Symp. Virtual Reality Software and Technology (VRST)*, pages 1–7, 2017.
[8] H. G. Kim, H.-T. Lim, S. Lee, and Y. M. Ro. VRSA net: VR sickness assessment considering exceptional motion for 360 VR video. *IEEE Trans. Image Process.*, 28(4):1646–1660, 2018.
[9] J. Kim, W. Kim, S. Ahn, J. Kim, and S. Lee. Virtual reality sickness predictor: Analysis of visual-vestibular conflict and VR contents. In *Int. Conf. Qual. Multi. Exp. (QoMEX)*, pages 1–6. IEEE, 2018.
[10] J. Kim, W. Kim, H. Oh, S. Lee, and S. Lee. A deep cybersickness predictor based on brain signal analysis for virtual reality contents. In *IEEE Proc. Int. Conf. Comput. Vis.(ICCV)*, pages 10580–10589, 2019.
[11] J. Kim, A.-D. Nguyen, and S. Lee. Deep cnn-based blind image quality predictor. *IEEE Transactions on Neural Networks and Learning Systems*, 30(1):11–24, 2018.
[12] J. Kim, H. Oh, W. Kim, S. Choi, W. Son, and S. Lee. A deep motion sickness predictor induced by visual stimuli in virtual reality. *IEEE Trans. Neural Netw. Learn. Syst.*, 2020.
[13] W. Kim, S. Ahn, A.-D. Nguyen, J. Kim, J. Kim, H. Oh, and S. Lee. Modern trends on quality of experience assessment and future work. *APSIPA Trans. Signal Inf. Process.*, 8, 2019.
[14] W. Kim, J. Kim, S. Ahn, J. Kim, and S. Lee. Deep video quality assessor: From spatio-temporal visual sensitivity to a convolutional neural aggregation network. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 219–234, 2018.
[15] W. Kim, S. Lee, and A. C. Bovik. VR sickness versus VR presence: A statistical prediction model. *IEEE Trans. Image Process.*, 30:559–571, 2020.
[16] W. Kim, A.-D. Nguyen, S. Lee, and A. C. Bovik. Dynamic receptive field generation for full-reference image quality assessment. *IEEE Trans. Image Process.*, 29:4219–4231, 2020.
[17] N. Padmanaban, T. Ruban, V. Sitzmann, A. M. Norcia, and G. Wetzstein. Towards a machine-learning approach for sickness prediction in 360 stereoscopic videos. *IEEE Trans. Vis. Comput. Graph.*, 24(4):1594–1603, 2018.
[18] J. T. Reason and J. J. Brand. *Motion sickness.* Academic press, 1975.
[19] S. Sharples, S. Cobb, A. Moody, and J. R. Wilson. Virtual reality induced symptoms and effects (VRISE): Comparison of head mounted display (HMD), desktop and projection display systems. *Displays*, 29(2):58–69, 2008.
[20] VQEG. Final report from the video quality experts group on the validation of objective models of video quality assessment, Mar. 2000.
[21] D. Walther, U. Rutishauser, C. Koch, and P. Perona. On the usefulness of attention for object recognition. In *Workshop on Attention and Performance in Computational Vision at ECCV*, pages 96–103, 2004.
[22] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240, 1998.