VRSA Net: VR Sickness Assessment Considering Exceptional Motion for 360° VR Video

Hak Gu Kim, *Student Member, IEEE*, Heoun-Taek Lim, Sangmin Lee, and Yong Man Ro, *Senior Member, IEEE*

***Abstract*— The viewing safety is one of the main issues in viewing virtual reality (VR) content. In particular, VR sickness could occur when watching immersive VR content. To deal with the viewing safety for VR content, objective assessment of VR sickness is of great importance. In this paper, we propose a novel objective VR sickness assessment (VRSA) network based on deep generative model for automatically predicting the VR sickness score. The proposed method takes into account motion patterns of VR videos in which an exceptional motion is a critical factor inducing excessive VR sickness in human motion perception. The proposed VRSA network consists of two parts, which are VR video generator and VR sickness score predictor. By training the VR video generator with common videos with non-exceptional motion, the generator learns the tolerance of VR sickness in human motion perception. As a result, the difference between the original and the generated videos by the VR video generator could represent exceptional motion of VR video causing VR sickness. In the VR sickness score predictor, the VR sickness score is predicted by projecting the difference between the original and the generated videos onto the subjective score space. For the evaluation of VR sickness assessment, we built a new dataset which consists of 360° videos (stimuli), corresponding physiological signals, and subjective questionnaires from subjective assessment experiments. Experimental results demonstrated that the proposed VRSA network achieved a high correlation with human perceptual score for VR sickness.**

***Index Terms*— VR sickness, deep learning, virtual reality, objective assessment, motion mismatch.**

I. Introduction

VIRTUAL reality (VR) contents such as 360-degree video can provide realistic and immersive viewing experience for viewers. While conventional 2D rectangle image has a limited field of view (FOV) at a fixed viewpoint, the 360-degree video provides unlimited FOV in all directions [1], [2]. Viewers can see wherever they want to see by selecting the specific portion of spherical images (called as *viewport*) with VR displays such as a head-mounted display (HMD). The development of the 360-degree cameras and VR displays has increased the interest and popularity of the VR content (e.g., 360-degree video).

As the growth of the VR content services, concerns on the viewing safety are considerably increasing in viewing VR content. Many studies reported various physical symptoms such as headache, focusing difficulty and dizziness during VR content viewing, which were caused by VR sickness [3], [4]. VR sickness, which is one of the bottlenecks for proliferation of VR market, could induce three major symptoms: 1) oculomotor symptoms including visual fatigue and focusing difficulty, 2) disorientation symptoms including dizziness and vertigo, and 3) nausea symptoms including salivation, sweating, and burping [3], [4]. Approximately 80% to 95% of viewers exposed to VR experience reported some level of VR sickness [5].

There are various determinants of VR sickness, such as excessive motion mismatch, a wide FOV [6]–[8], time lag [9], [10], etc. In particular, the excessive motion mismatch between what viewers' eyes are seeing (i.e., simulation motion of VR content) and what viewers' ears are feeling (i.e., physical motion of viewers) leads to a high degree of sensory conflicts between visual sensor and vestibular sensor (i.e., visual-vestibular conflict [11]) [12]–[14]. The visual-vestibular conflict is largely caused by the exceptional motion (e.g., exceptional acceleration and rapid turning) of content since the physical motion of viewer is relatively static. The exceptional motion leading to VR sickness means the exceeded acceleration and rapid turning [15], [16], such as racing and roller coaster. For example, when watching a 360-degree roller coaster video with a HMD, our visual sensor tells us that you move very fast. Whereas, our vestibular sensor tells us that you are not in motion actually. As a result, the discrepancy mainly leads to excessive VR sickness in human motion perception system. In particular, the exceptional motion in immersive VR content could exacerbate motion mismatch so that it is highly correlated to VR sickness.

To deal with that, a lot of time and effort have been devoted for the creation of viewing safe VR contents [17]. In addition, the viewing safety issue has been raised for user-generated VR contents as well. Therefore, it is essential to develop the objective VR sickness assessment (VRSA) that automatically predicts the degree of VR sickness in VR viewing.

Most of existing works focused on measuring physiological signals [18]–[21] or scoring subjective questionnaires [5], [22]–[24] through subjective assessment experiments in a virtual environment. The conventional objective VRSA approaches were very cumbersome due to physiological mea-

Manuscript received May 21, 2018; revised September 2, 2018 and October 20, 2018; accepted October 21, 2018. Date of publication November 12, 2018; date of current version November 28, 2018. This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No. 2017-0-00780, Development of VR sickness reduction technique for enhanced sensitivity broadcasting). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Lei Zhang. *(Corresponding author: Yong Man Ro.)*

The authors are with the Image and Video Systems Lab, School of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, South Korea (e-mail: hgkim0331@kaist.ac.kr; ingheoun@kaist.ac.kr; sangmin.lee@kaist.ac.kr; ymro@ee.kaist.ac.kr).

Digital Object Identifier 10.1109/TIP.2018.2880509

surements such as Electroencephalography (EEG) and galvanic skin response (GSR) and subjective questionnaires such as simulation sickness questionnaires (SSQ). The physiological measurement is generally vulnerable to various noises and subject's movement so that the existing approaches based on the physiological signals could be inaccurate. In addition, one of the challenges for developing an objective VRSA is small number of dataset with the ground truth of human perception for VR sickness. The aim of this paper is to propose a novel deep learning-based objective VRSA which automatically predicts the degree of VR sickness with only VR content. The main contributions of this paper are summarized as follows.

1) We present a novel deep generative model-based objective VR sickness assessment network (VRSA Net) considering exceptional motion pattern in VR content. In this paper, we propose a new objective VRSA framework, which consists of two parts, VR video generator and VR sickness score predictor. First, the VR video generator is trained with normal videos with non-exceptional motion in unsupervised manner. So the proposed generator is to learn the tolerance level of VR sickness in human motion perception. As a result, the generator well-reconstructs the VR videos with non-exceptional motion. On the other hand, the VR videos with exceptional motion patterns, which are highly likely to induce VR sickness, cannot be reconstructed well. The difference between the original video and generated video by the VR generator represents exceptional motion of VR video causing VR sickness. Second, the VR sickness score predictor is trained by mapping the difference between the original and the generated VR videos onto the corresponding subjective score. The VR sickness score can be assessed from the difference between the original and the generated videos. The combined architecture of 'generator (for normal videos) with unsupervised learning' and 'predictor with supervised learning' is our contribution for objective assessment.
2) For the evaluation of the proposed objective VRSA, we built a newly collected 360-degree video dataset with the corresponding subjective scores and physiological signal data, as a benchmark for VRSA. We collected 360-degree videos with different motion patterns as stimuli for our subjective VR experiments. The collected motion patterns are subjectively divided into three groups, which are slow, normal, complex motion pattern groups (see TABLE III). With the VR contents with different motion patterns, we conducted extensive subjective assessment experiments to verify the effectiveness of the proposed VRSA Net. We measured the level of VR sickness of subjects exposed to the VR contents using SSQ scores, heart rate and GSR. The prediction performance of the objective VRSA was evaluated with subjective SSQ scores (ground-truth) and physiological signals (heart rate and GSR). The dataset (i.e., VR contents and the corresponding SSQ scores and physiological signals) is publicly available on online [25].

Experimental results show that the proposed VRSA metric has a high correlation with the human subjective scores of in VRSA of VR viewing. In particular, substantial VRSA improvement (about 19 % increase of PLCC) can be achieved by the proposed method, compared to the assessment with physiological signals (heart rate and GSR). Furthermore, it is demonstrated that the proposed network not only measures the level of VR sickness, but also can detect which region mainly causes VR sickness.

The remainder of this paper is organized as follows: Section II briefly reviews the related works. Section III explains the proposed VRSA Net. Section IV describes the database used in the performance evaluation of VRSA. Specifically, we describe our subjective assessment experiments to obtain physiological signals (heart rate and galvanic skin conductance) and subjective questionnaires (SSQ) for VR sickness. In Section V, the performance of the proposed VRSA Net is evaluated. Finally, Section VI and VII provide discussions and conclusions, respectively.

II. Related Works

A. Image Quality Assessment of VR Content

Compared to the 2D rectangle image, the VR content such as 360-degree image has different characteristics including infinite field of view and projections from a spherical to a rectangle plane. The property of the VR content could cause distortion patterns such as rendering distortion and nonhomogeneous spatial distortion [26]–[28]. To deal with such characteristics, several studies of Image Quality Assessment (IQA) for VR content were reported. In [26], a spherical-based PSNR (S-PSNR) was proposed. It measured the quality of omnidirectional image by averaging the PSNR over the entire set of correspondences on the sphere. Sun *et al.* [27] proposed a weighted-to-spherically-uniform PSNR (WS-PSNR) method. They took into account the weights according to the pixel position on the spherical surface for accurately predicting the quality of the VR content. In [28], a Craster parabolic projection-based PSNR (CPP-PSNR) method was proposed in order to accurately measure the quality using Craster parabolic projection, which could reduce spatial distortion. In [29], a deep learning-based VR-IQA method was proposed, where an adversarial learning was employed so that the assessment performance of a degraded VR image could be improved.

B. VR Sickness Assessment

There were several studies of evaluating the VR sickness with subjective study and physiological measurement [18]–[21], [30]. In [18], the characteristic changes in physiology of cybersickness were investigated by measuring electrophysiological signals (EEG, electrogastrogram (EGG), GSR, etc.) of subjects exposed to VR contents. Based on the positive correlation between cybersickness by VR content (VR sickness) and physiological signals, the results showed that VR sickness accompanied the changes in the activity of the central and autonomic nervous systems. In [19], a subjective experiment was performed to measure various physiological signals during the virtual environment navigation with a HMD.

Experimental results provided that the changes in physiological signals such as eye blinking, stomach activity, and breathing could be caused by sensory mismatches between signals of real and virtual world. In [30], the quality of experience (QoE) and VR sickness of 360-degree videos were measured with mean opinion score (MOS) and SSQ, respectively. For a practical VR sickness evaluation on VR content, it is cumbersome to measure subjective questionnaires or physiological signals every time on subject viewing VR content. The proposed method in this paper predicts VR sickness based on VR content analysis without measuring cumbersome physiological signals or subjective questionnaires.

C. Deep Learning for Visual Quality Assessment

To deal with viewing safety, it is important to develop an objective assessment metric [31]–[35]. Recently, deep learning-based objective 2D and 3D IQA methods were proposed and provided the state-of-the-art prediction accuracy by modeling human visual perception [36]–[40]. In [36], a deep convolutional neural network (CNN)-based framework was proposed for full reference image quality assessment (FR-IQA), named as DeepQA. In [37], a novel multi-task end-to-end optimized deep network (MEON) was proposed using two sub networks for blind image quality assessment. In [38], a deep neural network-based IQA model was proposed for FR-IQA and no-reference image quality assessment (NR-IQA). Bosse *et al.* [38] devised a deep Siamese network model for FR-IQA and a CNN model for NR-IQA. In each model, by jointly learning the local weight and quality, the global image quality could be estimated accurately. In [39], a deep learning-based NR-IQA model for stereoscopic 3D (S3D) image was proposed. From the S3D images, the local features were extracted and aggregated to estimate the quality of S3D image by the CNN-based regression model [39]. In [40], a deep learning-based S3D visual comfort assessment (S3D-VCA) model was proposed considering the human attention model. Kim *et al.* [41] proposed a Binocular Fusion Net for S3D-VCA. In [41], by combining the spatial features of left and right views using a novel deep architecture, the latent binocular characteristics of stereoscopic images are learned to predict the visual comfort score in stereoscopic viewing. The existing deep learning-based models for predicting the image quality score or visual comfort scores were designed to regress the subjective score (i.e., ground-truth in training stage) from the high-level deep features. To train the deep learning model for human subjective score regression directly, it is necessary to collect large size databases including a lot of images and the corresponding subjective scores. In this paper, the proposed objective assessment framework for VRSA reliably predicts the level of VR sickness with a small scale of VR contents and the associated subjective scores through the two training processes. At first, the proposed VR video generator is trained with a large number of normal contents with tolerable factors in unsupervised manner. Based on the difference from the tolerable state (non-exceptional motion), then, VR sickness score predictor is trained to map the difference onto the VR sickness score.

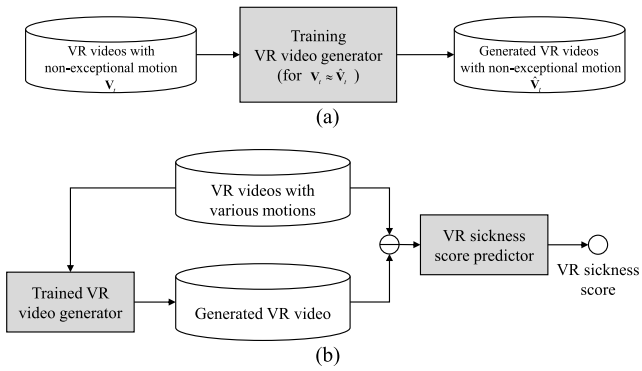


Fig. 1. Overall process of the proposed VRSA framework. (a) First, the VR video generator is trained with normal videos with non-exceptional motion in unsupervised manner. (b) Second, the VR sickness score predictor is trained with the difference between the original video and generated videos by the trained generator.

III. Proposed Method

A. Overview of the Proposed VR Sickness Assessment Framework

Figure 1 shows the overall process of the proposed VRSA framework in training. First, the VR video generator is trained with normal videos with non-exceptional motion in the manner of unsupervised learning. By learning the spatio-temporal characteristics of normal videos with non-exceptional motion pattern, the VR video generator is to learn a tolerance level of VR sickness in human motion perception. The trained VR video generator can well-reconstruct the VR videos with non-exceptional motion. On the other hand, the VR videos with exceptional motion pattern, which exceed the tolerance level of VR sickness in human motion perception, cannot be reconstructed well by the trained generator. So, the quality of the generated VR videos by the trained generator is correlated with exceptional motion causing VR sickness. After obtaining the generated VR videos by the trained generator, the VR sickness predictor is trained so that the differences between original and generated VR videos are regressed onto the ground-truth VR sickness score. The SSQ score plays a role of a ground-truth in order to train the predictor for VR sickness score prediction. A more detailed description of the proposed VRSA Net is described in the following subsections.

B. VR Video Generation for Learning the Tolerance of VR Sickness in Human Motion Perception

People usually experience non-exceptional motion in daily life but do not often experience exceptional motion. Therefore, human motion perception is tolerant of non-exceptional motion because non-exceptional motion could be well expected from the experience stored in neural store [42], but not tolerant of exceptional motion [43]. The video generator is trained with normal videos with non-exceptional motion so that it learns the tolerance level of VR sickness in human motion perception. Figure 2 shows examples of 360-degree videos with non-exceptional motion and exceptional motion patterns. The non-exceptional motion pattern of VR video is defined as slow and normal movement that people can see often in daily life such as stationary, walk, and normal driving.

(a) (b)

Fig. 2. Examples of 360-degree videos represented by equirectangular projection. (a) Examples of videos with non-exceptional motion such as scenery and normal driving (their SSQ scores < 28), (b) Examples of videos with exceptional motion such as racing and roller coaster (their SSQ scores > 40). Note that exceptional motion patterns inducing excessive VR sickness indicate acceleration and rapid turning [15].

It does not lead to severe VR sickness. On the other hand, the exceptional motion pattern of VR video is defined as acceleration and rapid turning [15], which are likely to cause severe VR sickness. Normal VR videos with non-exceptional motion do not cause VR sickness and their SSQ scores are under 28 [18], [44]. On the other hand, SSQ scores of the videos with exceptional motion (e.g., roller coaster and racing) are over 40 [18], [44]. In the training of the VR video generator, we use the normal video without exceptional motion [15], [16] for learning the tolerance of VR sickness in human motion perception.

Figure 3 shows the proposed VR video generator, which consists of the spatio-temporal generator for reconstruction of VR videos with non-exceptional motion and the spatio-temporal discriminator in the use of determining realistic videos with non-exceptional motion. By combining the spatio-temporal generator of spatio-temporal autoencoder (CNN+ConvLSTM) with spatio-temporal discriminator of 3D CNN in an adversarial way, in the proposed VR video generator, the video sequence is not only generated by the spatio-temporal autoencoder, it is also refined to be similar to the spatio-temporal characteristics of normal video with non-exceptional motion by the spatio-temporal discriminator. During the training, the spatio-temporal generator tries to reconstruct the video as much as the original. The spatio-temporal discriminator takes the original video or the generated video. Then, it determines whether a given video is realistic videos with non-exceptional motion or not. By adversarial learning between the generator and the discriminator, the proposed generator is able to synthesize the realistic video with non-exceptional motion.

1) Spatio-Temporal Generator for Reconstructing Normal VR Videos: In the proposed method, pleasantly-looking normal field-of-view (NFOV) segments from infinite FOV of 360-degree videos are used as input frames (i.e., spatio-temporal glimpse [45]), which are representative for the 360-degree video. To choose the NFOV, we first choose a center viewpoint in a form of longitude and latitude coordinates in the spherical domain. Then, a NFOV region is extracted from 360 degree-video frame by equirectangular projection

TABLE I

THE ARCHITECTURE OF THE SPATIO-TEMPORAL GENERATOR

Generator	Layer	Filter / Stride	Output size ($W \times H \times C$)
Spatial encoder	Conv 1, 2	3×3 / (1,1)	$224 \times 224 \times 64$
	Max pool	2×2 / (2,2)	$112 \times 112 \times 64$
	Conv 3, 4	3×3 / (1,1)	$112 \times 112 \times 128$
	Max pool	2×2 / (2,2)	$56 \times 56 \times 128$
	Conv 5, 6, 7	3×3 / (1,1)	$56 \times 56 \times 256$
	Max pool	2×2 / (2,2)	$28 \times 28 \times 256$
	Conv 8	3×3 / (1,1)	$28 \times 28 \times 512$
Temporal encoder	ConvLSTM 1	3×3 / (1,1)	$28 \times 28 \times 256$
Temporal decoder	ConvLSTM 2	3×3 / (1,1)	$28 \times 28 \times 512$
Spatial decoder	Deconv 1	3×3 / (2,2)	$28 \times 28 \times 256$
	Deconv 2, 3, 4	3×3 / (2,2)	$56 \times 56 \times 256$
	Deconv 5, 6	3×3 / (2,2)	$112 \times 112 \times 128$
	Deconv 7, 8	3×3 / (2,2)	$224 \times 224 \times 1$

with the viewpoint as a center. In this work, we set the size of an NFOV region to span 110-degree diagonal FOV, same as that of the mainstream VR headset. Let $\mathbf{I}_t$ and $\hat{\mathbf{I}}_t$ denote the t-th input frame and the t-th reconstructed frame, respectively. Let $\mathbf{V}_t$ and $\hat{\mathbf{V}}_t$ denote a set of original NFOV video frames (i.e., $\mathbf{V}_t = \left[\mathbf{I}_{t-N}, \cdots, \mathbf{I}_{t-1}, \mathbf{I}_t, \mathbf{I}_{t+1}, \cdots, \mathbf{I}_{t+N}\right]$) and a set of the generated NFOV video frames (i.e., $\hat{\mathbf{V}}_t = \left[\hat{\mathbf{I}}_{t-N}, \cdots, \hat{\mathbf{I}}_{t-1}, \hat{\mathbf{I}}_t, \hat{\mathbf{I}}_{t+1}, \cdots, \hat{\mathbf{I}}_{t+N}\right]$), respectively. As shown in Fig. 3, the proposed spatio-temporal generator consists of spatial encoder/decoder and temporal encoder/decoder. For spatial encoder and decoder, VGG-16 and "upside down" VGG-16 networks are employed, respectively [46], [47]. For temporal encoder and decoder, a convolutional LSTM (ConvLSTM) is employed [48]. TABLE I shows the architecture of the spatio-temporal generator of our VR video generator. In the spatial encoder, a spatial feature is encoded to represent visual characteristic of each frame. In this paper, the feature map of 8-th convolution layer of VGG-16 is used as the spatial feature denoted by $\mathbf{f}_t^{c8} \in \Re^{28 \times 28 \times 512}$. To learn the spatio-temporal feature, then, the spatial feature, $\mathbf{f}_t^{c8}$ is fed into the ConvLSTM. Let $\mathbf{h}_t^1 \in \Re^{28 \times 28 \times 256}$ and $\mathbf{h}_t^2 \in \Re^{28 \times 28 \times 512}$ denote the hidden states of ConvLSTM at l-th layer ($l = 1, 2$). In the temporal encoder and decoder, temporal characteristics of the training video dataset are learned. Finally, the original video sequence is reconstructed from the learned spatio-temporal features by the spatial decoder, "upside down" VGG-16 [47]. The t-th reconstructed frame $\hat{\mathbf{I}}_t$ can be represented by

$$\hat{\mathbf{I}}_t = G_\theta\left(\mathbf{I}_t\right) = \sigma_{dec}\left(\mathbf{W}_{dec}\mathbf{h}_t^2 + \mathbf{b}_{dec}\right), \tag{1}$$

where G_θ indicates the spatio-temporal generator with parameters θ. $\mathbf{W}_{dec}$ and $\mathbf{b}_{dec}$ represent the weight matrix and the bias vector of the spatial decoder, respectively. σ_{dec} is activation function of the spatial decoder.

Through adversarial learning, the generator reconstructs the video sequence (i.e., spatio-temporal glimpse) containing normal motion pattern (i.e., non-exceptional motion) well so that it attempts to deceive the discriminator (see Section III-B.2). To that end, the loss function of our generator is composed

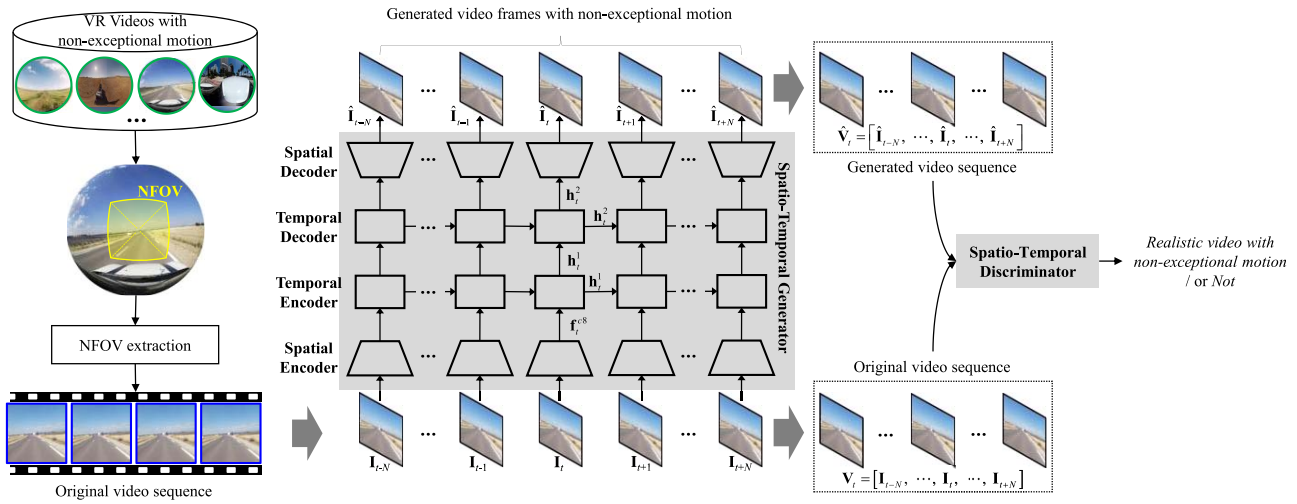


Fig. 3. The architecture of the proposed VR video generator for learning the tolerance of VR sickness in human motion perception. The proposed generator is trained with normal videos with non-exceptional motion. The proposed VR video generator consists of the spatio-temporal generator and discriminator with adversarial learning.

of two terms, which are realism loss, l_{real}, and reconstruction loss, l_{recon}. The realism loss can be written as

$$l_{real}(\theta; t) = -\log\left(D_\phi\left(G_\theta\left(\mathbf{I}_t\right)\right)\right), \tag{2}$$

where D_ϕ indicates the discriminator with parameters ϕ. By minimizing the realism loss, Eq. (2), the video generator forces the discriminator to consider the video generated by G_θ, $\hat{\mathbf{V}}_t$, as the original video, $\mathbf{V}_t$.

The reconstruction loss between the original and the generated frames, l_{recon}, can be written as

$$l_{recon}(\theta; t) = \frac{1}{2N+1}\sum_{k=t-N}^{t+N} \|G_\theta(\mathbf{I}_k) - \mathbf{I}_k\|_2^2. \tag{3}$$

By minimizing the reconstruction loss between the original frame $\mathbf{I}_t$ and the generated frame $\hat{\mathbf{I}}_t$, the reconstruction quality of the video with non-exceptional motion pattern can be enhanced. Finally, the total loss of the proposed spatio-temporal generator can be defined as a combination of the realism loss and the reconstruction loss.

$$L_G(\theta) = l_{real}(\theta; t) + \lambda_g l_{recon}(\theta; t), \tag{4}$$

where λ_g is a weight parameter to control the balance between the realism loss and reconstruction loss.

2) Spatio-Temporal Discriminator for Determining Realistic Normal VR Video: To improve the reconstruction performance of videos with non-exceptional motion during training, we devise the spatio-temporal discriminator with adversarial learning. The proposed discriminator is to determine whether the input video is a realistic video with non-exceptional motion or not by considering its spatio-temporal characteristics. As shown in Fig. 3, in training, the discriminator takes original video, $\mathbf{V}_t$ or the generated video, $\hat{\mathbf{V}}_t$. Then, it produces 1×1 output value in order to decide original video or the generated video. As seen in TABLE II, the proposed spatio-temporal discriminator is based on the 3D CNN structure. Our discriminator loss, L_D, can be written as

$$L_D(\phi) = \log\left(1 - D_\phi\left(\hat{\mathbf{V}}_t\right)\right) + \log\left(D_\phi\left(\mathbf{V}_t\right)\right). \tag{5}$$

TABLE II

THE ARCHITECTURE OF THE SPATIO-TEMPORAL DISCRIMINATOR

Layer	Filter / Stride	Output size ($D \times W \times H \times C$)
3D Conv 1	$5 \times 5 \times 5$ / (1,2,2)	$7 \times 112 \times 112 \times 32$
3D Conv 2	$3 \times 5 \times 5$ / (1,2,2)	$5 \times 56 \times 56 \times 64$
3D Conv 3	$3 \times 3 \times 3$ / (1,2,2)	$3 \times 28 \times 28 \times 128$
3D Conv 4	$3 \times 3 \times 3$ / (1,2,2)	$1 \times 14 \times 14 \times 256$
3D Conv 5	$1 \times 3 \times 3$ / (1,2,2)	$1 \times 7 \times 7 \times 1$
64-d fc layer	-	1×1

In Eq. (5), the $D_\phi(\hat{\mathbf{V}}_t)$ in the first term is the probability that the discriminator determines the generated video as original video. The second term in Eq. (5), $D_\phi(\mathbf{V}_t)$, is the probability that the discriminator determines the original video as original.

The proposed generator G_θ and discriminator D_ϕ form a generative adversarial network (GAN) [49]. To well learn the non-exceptional motion pattern with the adversarial learning, we learn the proposed VR video generator based on GAN. To well reconstruct the VR video with non-exceptional motion, the loss of the generator, L_G, is minimized. Alternatively, the loss of the discriminator, L_D, is maximized to precisely determine the realistic video with non-exceptional motion or not. Let λ_D denote a weight parameter of L_D for balance between L_G and L_D.

By performing the adversarial learning between G_θ and D_ϕ, the reconstruction performance of the generator and the discrimination performance of the discriminator can be improved together.

C. VR Sickness Score Predictor

Figure 4 shows the proposed VR sickness score predictor for automatically measuring the VR sickness score based on the difference between the original and the generated videos. By mapping the difference from the tolerance of human perception onto subjective sickness score, the proposed network architecture can reliably predict the subjective score

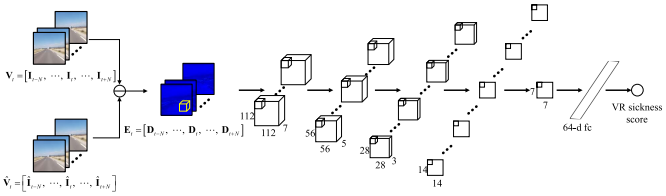


Fig. 4. The architecture of the proposed VR sickness score predictor for measuring the VR sickness score.

with a small scale of VR sickness datasets, compared to the conventional network architecture (e.g., 3D CNN) to directly assess the subjective score from contents. As shown in Fig. 4, after obtaining the generated videos by the trained VR video generator, the difference between the original and the generated videos ($\mathbf{D}_t$) is calculated as

$$\mathbf{D}_t = |\mathbf{I}_t - G_\theta(\mathbf{I}_t)| = \left|\mathbf{I}_t - \hat{\mathbf{I}}_t\right|. \tag{6}$$

Since the VR video generator is trained with normal videos with non-exceptional motion, i.e., tolerable data for VR sickness in human motion perception, the reconstruction error of the generated video could represent exceptional motion of VR video causing VR sickness. Thus, the $\mathbf{D}_t$ indicates the distance from the tolerance of VR sickness in the human motion perception. Based on the difference between the original and the generated video frames, $\mathbf{D}_t$, VR sickness score predictor quantifies the level of VR sickness. Let $\mathbf{E}_t$ denote a set of the difference maps at each frame (i.e., $\mathbf{E}_t = [\mathbf{D}_{t-N}, \cdots, \mathbf{D}_{t-1}, \mathbf{D}_t, \mathbf{D}_{t+1}, \cdots, \mathbf{D}_{t+N}]$). As shown in Fig. 4, the VR sickness score predictor consists of 3D CNN for encoding the spatio-temporal characteristics of a sequence of the difference maps, $\mathbf{E}_t$. The architecture of the predictor is the same as that of the discriminator (see TABLE II). In the VR sickness score predictor part, the difference map from the VR generator learning normal motion perception is mapped to the human subjective score space. In training, the VR sickness score is predicted from $\mathbf{E}_t$ by minimizing the loss between the predicted score and ground-truth VR sickness score. In this paper, the total SSQ score obtained from subjective assessment experiment is used as ground-truth subjective score in training (see Section IV). The loss function of the VR sickness score predictor, L_P, is defined as

$$L_P = \frac{1}{K}\sum_{k=1}^{K}\left\| f_{VRS}\left(\mathbf{E}_t^k\right) - SSQ_{total}^k \right\|^2, \tag{7}$$

where K is the number of batches and SSQ_{total}^k denotes the ground-truth subjective score, which is the total SSQ score of k-th video sequence. $f_{VRS}(\cdot)$ is the function of the VR sickness score predictor using 3D CNN. $f_{VRS}\left(\mathbf{E}_t^k\right)$ indicates the predicted VR sickness score.

In training, the parameters of the predictor are trained by minimizing the loss, Eq. (7). In testing, at first, the difference maps are obtained by the trained VR video generator. With the sequence of the difference maps $\mathbf{E}_t$, then, the final VR sickness score is yielded by the trained VR sickness score predictor.

IV. Benchmark Database for VR Sickness Evaluation

For evaluation of the prediction performance of the objective VRSA, it is required to obtain subjective VR sickness scores such as SSQ for stimuli of VR contents. The prediction performance of the objective VRSA methods can be evaluated by measuring the correlation between the subjective VR sickness scores and the predicted scores obtained from objective VRSA. In this paper, we built a newly collected 360-degree video database, the corresponding subjective scores and physiological signal data. In this section, we present the overall procedure of our subjective assessment experiments for collecting subjective scores and the physiological signals in watching 360-degree videos with a HMD.

A. 360-Degree Video Dataset

In this paper, we collected a 360-degree video dataset with high spatial resolution and various motion patterns. A few 360-degree image datasets have been recently introduced for saliency detection [50], [51]. In most of the previous works for subjectively and objectively assessing VR sickness, only one or two VR contents were used. We newly collected nine 360-degree videos including various scenes from Youtube as a benchmark and conducted an extensive subjective experiment for evaluation of objective VRSA. The videos contain various scenes such as beach, driving, flight, roller coaster, etc. To investigate the effect of various motion patterns on VR sickness, we collected 360-degree video datasets with various motion patterns from static to dynamic. They were subjectively divided into three categories based on motion patterns: simple, normal, and complex. The most of the collected contents have 4K resolution (3840×1920 or 3840×2048) with 30 Hz. Due to the viewing safety issue of the participated subjects, each test video was presented for 90 seconds (see Section IV-B).

TABLE III shows a detailed description for the dataset. In TABLE III, the motion pattern indicates the motion type of each video. "Slow" represents static and slow movement. "Normal" represents the movement of normal speed such as driving a car. "Complex" indicates fast acceleration and rapid rotations such as roller coaster, i.e., non-exceptional motion. The name and Youtube ID represent video name and Youtube video identification, respectively. With the video name and Youtube ID, we can access the dataset on the Web. Resolution in TABLE III indicates the spatial resolution of each video used in our experiment. FPS means frames per second of each video used in our experiment. Time stamp provides which part of the video has been played in our subjective experiment because each video was presented only for 90 seconds (not entire sequence) in our experiments for viewing safety of subjects.

B. Subjective Assessment Experiment With 360-Degree Video

In subjective assessment experiments, Oculus Rift CV1 was used for displaying 360-degree videos, which was one of the high-end stereoscopic type HMDs. Its display resolution is

TABLE III
DETAILED DESCRIPTION OF VR VIDEO DATASET

Motion pattern	No.	Name	Youtube ID	Resolution	Fps	Time stamp
Simple	1	Surrounded by wild elephants	mlOiXMvMaZo	3840 × 2048	30.00	01:00 - 02:30
	2	Maldives VR 360 - 4K video	MgJITGvVfR0&list=	3840 × 1920	29.97	00:10 - 01:40
	3	Cockpit-Flug in 360°	g-CxpFuiU9Q	3840 × 2048	29.97	08:00 - 09:30
Normal	4	Wingsuit 360-degree video	AX4hWfyHr5g	3840 × 1920	30.00	00:00 - 01:30
	5	Driving around Lavender fields	JEr3-FzSgzk	3840 × 1920	29.97	00:00 - 01:30
	6	Driving Winter Alpine Road	y6x2mc3xLX4	3840 × 1920	29.97	03:30 - 05:00
Complex	7	Superman Roller Coaster 360 VR	jLtcPTm5dTg	2560 × 1440	25.00	00:45 - 02:15
	8	Rallying in 360 with Peer-Thru	R1UMjiQ6AuU	5120 × 2560	59.94	00:50 - 02:20
	9	Jet Speed 360 VR 4K	lrgFxrYafDA	3840 × 2160	29.97	00:45 - 02:15

Note that the readers can access the dataset using Youtube ID of each video by entering the Youtube URL with Youtube ID: "https://www.youtube.com/watch?v="+"Youtube ID".

2160×1200 pixels (1080×1200 pixels per eye). Its display frame rate is maximum 90 Hz and it has 110 degree FoV.

A total of twenty subjects, aged 20 to 30, participated in our subjective experiments under the approval of KAIST Institutional Review Board (IRB). In general, the use of VR is not recommended for young people under 12 years of age due to immature development of visual-vestibular sensors. According to [52] and [53], older people reported more severe VR sickness due to the age related changes in the oculomotor system. The participants in our experiment do not have health problems such as immature development of visual-vestibular sensors, vestibular dysfunction or oculomotor dysfunction, compared to children and older people. Note that ITU-R BT.500-13 recommended at least fifteen subjects in order to obtain reliable subjective experiment results [54]. Subjects have normal or corrected-to-normal vision and minimum stereopsis of 60 arcsec. In our experiment, before watching each stimulus, they were placed in the center position to be started from zero position in order to prevent significantly different viewing traces between viewers [30]. They were seated on a rotatable chair in order to freely look around 360-degree contents. A week before the actual subjective assessment experiments, we had subjects experience a variety of VR contents with Oculus Rift in order to allow them to familiarize with VR environment. In our experiments, the subject head motion was small and negligible during watching 360-degree contents. Since most of the 360 degree-videos used in our experiment have movement in a certain direction by roller coaster and car, subjects focused their gaze in the similar direction (e.g., the direction of rails in the roller coaster video or moving direction in the driving video) [51]. The head motion below the range of 44° to 55° in yaw could not cause severe VR sickness [24]. All experimental environments followed the guideline as per the recommendations of ITU-R BT.500-13 [54] and BT.2021 [55].

Each test video was presented for 90 seconds. The order of presentation of each video was randomized across subjects. Then, resting time was given as 150 seconds with mid gray image. During the resting time, subjects were asked to assess the degree of perceived VR sickness. To grade the degree of VR sickness, the latest version of 16-item SSQ [56] was used in our experiment. The 16-item SSQ consists of 16 physical symptoms, which are highly related to VR sickness,

TABLE IV
16-ITEM SSQ USED IN OUR SUBJECTIVE ASSESSMENT FOR VR SICKNESS

No.	SSQ symptoms	Nausea	Oculomotor	Disorientation
1	General discomfort	O	O	
2	Fatigue		O	
3	Headache		O	
4	Eye strain		O	
5	Difficulty focusing		O	O
6	Increased salivation	O		
7	Sweating	O		
8	Nausea	O		O
9	Difficulty concentrating	O	O	
10	Fullness of head			O
11	Blurred vision		O	O
12	Dizzy (Eyes open)			O
13	Dizzy (Eyes closed)			O
14	Vertigo			O
15	Stomach awareness	O		
16	Burping	O		

with a discrete four point grading scale for each symptom (0: None, 1: Slight, 2: Moderate, 3: Severe). TABLE IV shows the16-item SSQ. The SSQ scores of three major symptoms are calculated by summation of scores for each symptom included in their categories with weight: 9.54 for nausea, 7.58 for oculomotor, 13.92 for disorientation, respectively [3], [18], [56]. The SSQ score for nausea can be written as

$$SSQ_{Nausea} = 9.54 \times \frac{1}{J} \sum_{j=1}^{J} \left(s_j^{gd} + s_j^{is} + s_j^{s} + s_j^{n} + s_j^{dc} + s_j^{sa} + s_j^{b} \right), \quad (8)$$

where J is the number of subjects. s_j^{gd}, s_j^{is}, and s_j^{s} are subjective scores of j-th subject for general discomfort, increased salivation, and sweating symptoms, respectively. s_j^{n}, s_j^{dc}, s_j^{sa}, and s_j^{b} are subjective scores of j-th subject for nausea, difficulty concentrating, stomach awareness and burping, respectively.

The SSQ score for oculomotor can be written as

$$SSQ_{Oculo} = 7.58 \times \frac{1}{J} \times \sum_{j=1}^{J} \left(s_j^{gd} + s_j^{f} + s_j^{h} + s_j^{es} + s_j^{df} + s_j^{dc} + s_j^{bv} \right), \quad (9)$$

where s_j^f, s_j^h, and s_j^{es} are subjective scores of j-th subject for fatigue, headache, and eye strain, respectively. s_j^{df} and s_j^{bv} are subjective scores for difficulty focusing and blurred vision, respectively.

The SSQ score for disorientation can be written as

$$SSQ_{Dis} = 13.92 \times \frac{1}{J} \times \sum_{j=1}^{J} \left(s_j^{dj} + s_j^{n} + s_j^{fh} + s_j^{bv} + s_j^{dzo} + s_j^{dzc} + s_j^{v} \right), \quad (10)$$

where s_j^{fh}, s_j^{dzo}, s_j^{dzc}, and s_j^{v} are subjective scores of j-th subject for fullness of head, dizzy (eye open), dizzy (eye closed), and vertigo, respectively.

Finally, a total SSQ score was obtained by combining the partial SSQ scores for three major symptoms with the weight, 3.74 [3], [18], [56], which can be written as

$$SSQ_{total} = 3.74 \times \left(\frac{1}{9.54} SSQ_{Nausea} + \frac{1}{7.58} \times SSQ_{Oculo} + \frac{1}{13.92} SSQ_{Dis} \right). \quad (11)$$

At the same time, we measured skin conductance and heart rate of subjects during our subjective assessment for objective evaluation of VR sickness. Heart rate and skin conductance were measured using NeuLog heart rate/pulse sensor (NUL-208) and GSR sensor (NUL-207) for measuring the physiological signals in watching VR contents to build the benchmark database. The heart rate/pulse sensor was composed of an infrared LED transmitter and a matched infrared phototransistor receiver. The GSR sensor was composed of two probes and finger connectors. Their maximum sampling rate was 100 Hz. To obtain baseline signals of each subject, after the subjects comfortably relaxed for 5 minutes, we measured the baseline physiological signals during same period (90 seconds) of the video viewing before the subjective assessment. In our experiment, to eliminate the sickness caused by continuously watching VR content, before presenting next VR content, we asked subjects to tell about the current degree of VR sickness on a scale of 0 – 20 using fast motion sickness scale (FMS) [57]. When they told 0 score (no sickness), we continuously conducted the experiment. Otherwise, we gave the subject additional resting time until they told 0 score for VR sickness. The additional resting time for each stimulus was about 60 sec averagely. As a result, total resting time for each stimulus was about 210 sec (150 sec for basic resting time + 60 sec for additional resting time), which was more than twice the presentation time of each test video. As such, each subject took about 60 min to complete the subjective assessments including time to attach the equipment. During the subjective assessment experiment, the subjects were allowed to immediately stop and take a break if they feel difficult to continue the experiment due to excessive VR sickness.

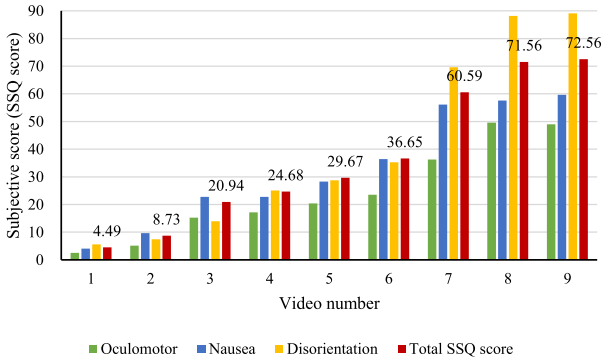


Fig. 5. Subjective assessment results of VR sickness for the 360-degree video dataset. The x-axis and y-axis represent the video data number and SSQ scores for VR sickness. Note that the green and blue bars represent the SSQ scores for oculomotor and nausea, respectively. The yellow and red bars represent SSQ scores for disorientation and total SSQ score, respectively. The total SSQ score for each test video is labeled over the red bar.

C. Subjective Assessment Results

Figure 5 shows the SSQ scores for the VR sickness of collected 360-degree video datasets. The x-axis and y-axis indicate the video number and corresponding SSQ scores for VR sickness, respectively. In Fig. 5, green and blue bars represent SSQ scores for oculomotor and nausea, respectively. The yellow and red bars represent SSQ scores for disorientation and total SSQ score, respectively. The total SSQ score for each test video is labeled over the red bar. As shown in Fig. 5, the total SSQ scores of the 360-degree videos with slow and normal motion patterns (i.e., video data number: 1 to 6) were low. Note that the total SSQ scores ranging of 32 to 40 indicate noticeable VR sickness [18]. It means that the viewers could not perceive severe VR sickness in watching VR contents with slow or normal motion pattern since the discrepancy between the simulation motion and physical motion is not excessive. On the other hand, the total SSQ scores of the VR contents with complex motion pattern such as roller-coaster (total SSQ: 60.59), rally racing (total SSQ: 71.56), and jet racing (total SSQ: 72.56) were much higher than those of VR contents with slow and normal motion patterns. In particular, in the SSQ scores of VR video data number 7, 8, and 9, the SSQ scores for disorientation were generally higher than those of oculomotor and nausea. The result indicates that the complex motion patterns (i.e., exceptionally fast and rotational motions) could have mainly an influence on the disorientation factor.

In this paper, the benchmark has corresponding subjective sickness scores and nine 360-degree videos (in Section IV-A) which have various motion patterns (e.g., slow, normal, and complex) and scenes (e.g., scenery, driving, flight, roller coaster, racing, etc.). Regarding non-exceptional motion and exceptional motion, the benchmark with nine 360-degree videos and the corresponding subjective sickness scores is enough to perform the experiments compared with [8], [18], [20], [23], and [30] for evaluating VR sickness.

V. Experiments and Results

A. Experimental Setup and Deep Network Training

To verify the performance of the proposed VR sickness assessment deep network, experiments were conducted with the benchmark database that consists of the 360-degree video dataset and the corresponding SSQ scores and physiological signals (heart rate and skin conductance) of VR sickness. The experiments were conducted on a PC with Intel Core-4770 CPU @ 3.40 GHz, a 32 GBytes memory, and NVIDIA GTX 1080 TI. The proposed VRSA framework was implemented using TensorFlow.

For training VR video generator in the proposed VRSA framework, we used other various video datasets in the experiment, which are KITTI benchmark datasets [58] and various other 360-degree video contents from Vimeo. In the experiment, KITTI benchmark datasets were used for pre-training of our VR video generator. The KITTI benchmark database includes a total of 61 normal driving video clips with a resolution of 1242 × 375 pixels [58]. The normal driving clips have three types of scenes, which are city, residential, and road [58]. The number of frames in KITTI benchmark dataset for video generator training is 42,746 frames (42,746 frames = 8,477 frames from city clips + 28,404 frames from residential clips + 5,865 frames from road clips) [58]. From Vimeo, twenty 360-degree video clips with non-exceptional videos were collected (see Appendix A for more detail of the Vimeo twenty 360-degree video clips), which were used for training of the generator. A total of 18,000 frames in Vimeo dataset were used for the training (i.e., 900 frames × the twenty Vimeo 360-degree videos). The 900 frames in each video were chosen by selecting all frames in the 'time stamp' with 'fps' of each video in TABLE VII (The number of frames = 'fps' × 'time stamp'). As a result, a total of 60,746 frames (60,746 frames = 18,000 frames from Vimeo + 42,746 frames from KITTI benchmark dataset) were used in VR video generator training.

Unlike the video generator training in unsupervised manner, since videos and the corresponding subjective scores are required as a ground-truth for VR sickness predictor training, the VR sickness score predictor in the proposed VRSA framework was trained by another twenty one 360-degree videos which were captured by photo experts and available from [59] (see TABLE VIII in Appendix B for more detail of the twenty one 360-degree videos). Most of 360-degree videos had 3840 × 1920 or 4096 × 2048 pixels. A total of 56,700 frames were used for the training (i.e., 2,700 frames × the twenty one 360-degree videos). To obtain the subjective scores of the twenty one videos in TABLE VIII, we conducted additional subjective assessment experiment with other twenty subjects by the same methodology of the subjective experiments in Section IV-B. Figure 6 shows the total SSQ scores of the twenty one 360-degree videos (i.e., subjective assessment results), which were used as ground truths for training of the VR sickness score predictor.

To train the proposed VRSA network with VR video generator and VR sickness score predictor, two-step training was used. In the first step, for the training of the VR video generator, the VR video generator was pre-trained with KITTI dataset for the reconstruction of the videos with normal driving (non-exceptional motion). The pre-trained weights were used as initial parameters in training with the twenty 360-degree videos, which were collected from Vimeo. The VR video generator was trained again with the twenty 360-degree videos consisting non-exceptional motion in unsupervised manner (without SSQ scores). In the second step, the trained generator in the first step generated the generated videos for all input of twenty one 360-degree videos that consist of various motion patterns, for the training of the VR sickness score predictor. Then, a sequence of the difference maps, $\mathbf{E}_t$, was obtained by taking the difference between the original and the generated videos. With the $\mathbf{E}_t$ and the corresponding total SSQ score (see Fig. 6), the VR sickness score predictor was trained.

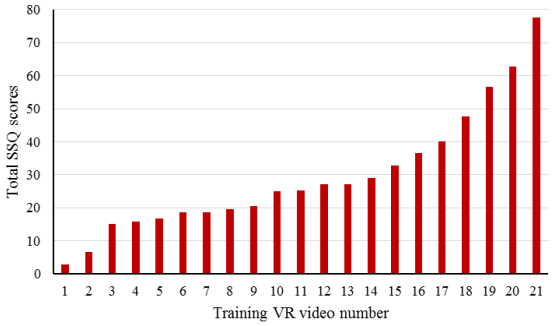


Fig. 6. The total SSQ scores of the twenty one 360-degree videos for training of the VR sickness score predictor.

The proposed VR video generator was pre-trained by 60 epochs with ADAM optimizer [60]. In each iteration, we used a batch size of 3. For ADAM optimizer, the learning rate was initialized at 0.00005. β_1 and β_2 were set to 0.9 and 0.999, respectively. Weight decay was set to 10^{-8} per each iteration. In training of the proposed VR video generator, at first, only the spatio-temporal generator was trained to minimize Eq. (4), L_G. Then, the spatio-temporal generator and the discriminator were alternately trained in an adversarial way. The VR sickness score predictor was trained with the same optimizer, learning rate, and batch size. At the end of the discriminator and the predictor, the sigmoid was used as an activation function. In our experiment, λ_D and λ_g were set to 1.

B. Prediction Performance Evaluation

For performance evaluation of the proposed VRSA, we used the benchmark database in Section IV that consists of nine 360-degree videos (TABLE III) and the associated SSQ scores (Fig. 5) and physiological signals (heart rate and skin conductance) of VR sickness. To evaluate the proposed objective VRSA method, we employed commonly used performance measures: Pearson linear correlation coefficient (PLCC), Spearman rank order correlation coefficient (SROCC), and root mean square error (RMSE).

To evaluate the prediction performance of the proposed method, we compared the performance with three physiological signal-based methods [18]–[20] (were heart rate, heart rate

TABLE V

PREDICTION PERFORMANCE OF THE PROPOSED METHOD AND OTHER METHODS

Objective metrics	PLCC	SROCC	RMSE
HR-based method	0.215	0.217	26.538
HRV-based method	0.515	0.400	21.172
GSR-based method	0.691	0.695	20.405
Optical flow-based method	0.717	0.715	16.592
Deep learning method	0.613	0.569	20.898
Proposed VRSA Net	**0.885**	**0.882**	**10.251**

TABLE VI

STATISTICAL ANALYSIS OF PREDICTION PERFORMANCES FOR DIFFERENT METRICS

	HR-based method	HRV-based method	GSR-based method	Optical flow-based method	Deep learning-based method	Proposed VRSA Net
HR-based method		110	110	111	110	111
HRV-based method	110		110	110	110	111
GSR-based method	110	110		000	110	111
Optical flow-based method	111	110	000		110	111
Deep learning-based method	110	110	110	110		111
Proposed VRSA Net	111	111	111	111	111	

Each entry corresponds to the results of the statistical test on the performance indexes (from left to right: PLCC, SROCC, and RMSE). "1" means that the difference between the prediction performances of two different metrics is statistically significant in 95% significance level. "0" means that the difference is not significant.

variability (HRV), and GSR). For physiological signal-based methods, the HR and GSR in the benchmark (see Section IV) were used. In the HR-based VRSA, the mean of heart rate in time domain value was used as objective metric using heart rate [18], [20]. For the HRV-based VRSA, the standard deviation of the heart rate in time domain was calculated from the heart rate signals [18]. In the GSR-based VRSA, each normalized GSR signal was obtained by subtracting the average GSR signal of test image from the average baseline GSR signal of subject. The mean of normalized GSR in time domain was used as objective metric using GSR [19]. In addition we performed VR sickness assessment by measuring the optical flow (motion information) of 360-degree videos. The average magnitude of optical flow was used as VR sickness metric. For performance comparisons, the performance metrics using physiological signals and optical flow were computed after nonlinear regression using logistic function [61]. For performance comparisons of deep learning-based approach, 3D CNN, which is one of the main architectures of deep learning for video analysis, was employed. The architecture of 3D CNN is the same architecture as in TABLE II). It consists of five 3D convolutional layers and 64-dimensional fully-connected layer. Similar to other deep learning-based objective assessment approaches [37], [38], [40], the deep learning-based method using 3D CNN was end-to-end trained with the dataset, which was used in the training of the proposed VR sickness predictor (twenty one 360-degree videos from [59] in TABLE VIII and the corresponding sickness scores obtained by our subjective experiment), in supervised manner.

TABLE V shows the results of the prediction performance evaluation for the proposed VRSA metric, three physiological signals-based methods, optical flow-based method, and deep learning-based method using 3D CNN. As seen in TABLE V, the results reveal that the proposed VRSA model yields a high correlation with subjective VR sickness score, i.e., total SSQ of the test datasets (PLCC: 0.885 and SROCC: 0.882). The RMSE value of the proposed method was significantly lower than those of the existing objective VRSA methods using physiological signals.

The proposed VRSA Net even without the predictor (i.e., the average difference value between the original and the generated videos are used as objective metric for VR sickness) achieved better prediction performance (PLCC: 0.869 and SROCC: 0.877) than other methods. The proposed VRSA Net with the predictor provided higher performance of about 2%, compared to the VRSA Net without the predictor. The results demonstrate that the proposed VR video generator can effectively capture the exceptional motion leading to excessive VR sickness and the predictor can precisely assess the VR sickness score by considering the total SSQ scores as references for the performance improvement. In the physiological signals-based VRSA methods, the HRV-based and GSR-based methods had a correlation with subjective VR sickness score and a total SSQ score. On the other hand, the heart rate did not seem to correlate with the degree of VR sickness. These results are consistent with [18]. Compared to the objective VRSA methods based on physiological measurements and deep learning-based method, the proposed VRSA Net achieved superior prediction performance. Importantly, these results indicate that motion mismatch caused by exceptional motion of VR content is one of the most important factors on the VR sickness.

In addition, the statistical significance evaluation was performed under the recommendation of ITU-T P.1401 [62]. The guideline provides the statistical evaluation and qualification procedure of the objective assessment models (Z-test for PLCC and SROCC, and F-test for RMSE). To see whether the difference in prediction performance between different metrics is statistically significant or not, we conducted statistical significance evaluation. As seen in TABLE VI, the difference between the HR-based and GSR-based methods was statistically significant. The difference between the GSR-based and the optical flow-based methods was not statistically significant in terms of PLCC, SROCC, and RMSE. On the contrary, the differences between the proposed VRSA Net and other methods were statistically significant in terms of PLCC, SROCC, and RMSE. It means that the proposed VRSA method could be useful for VR sickness prediction without cumbersome measurement of physiological signals.

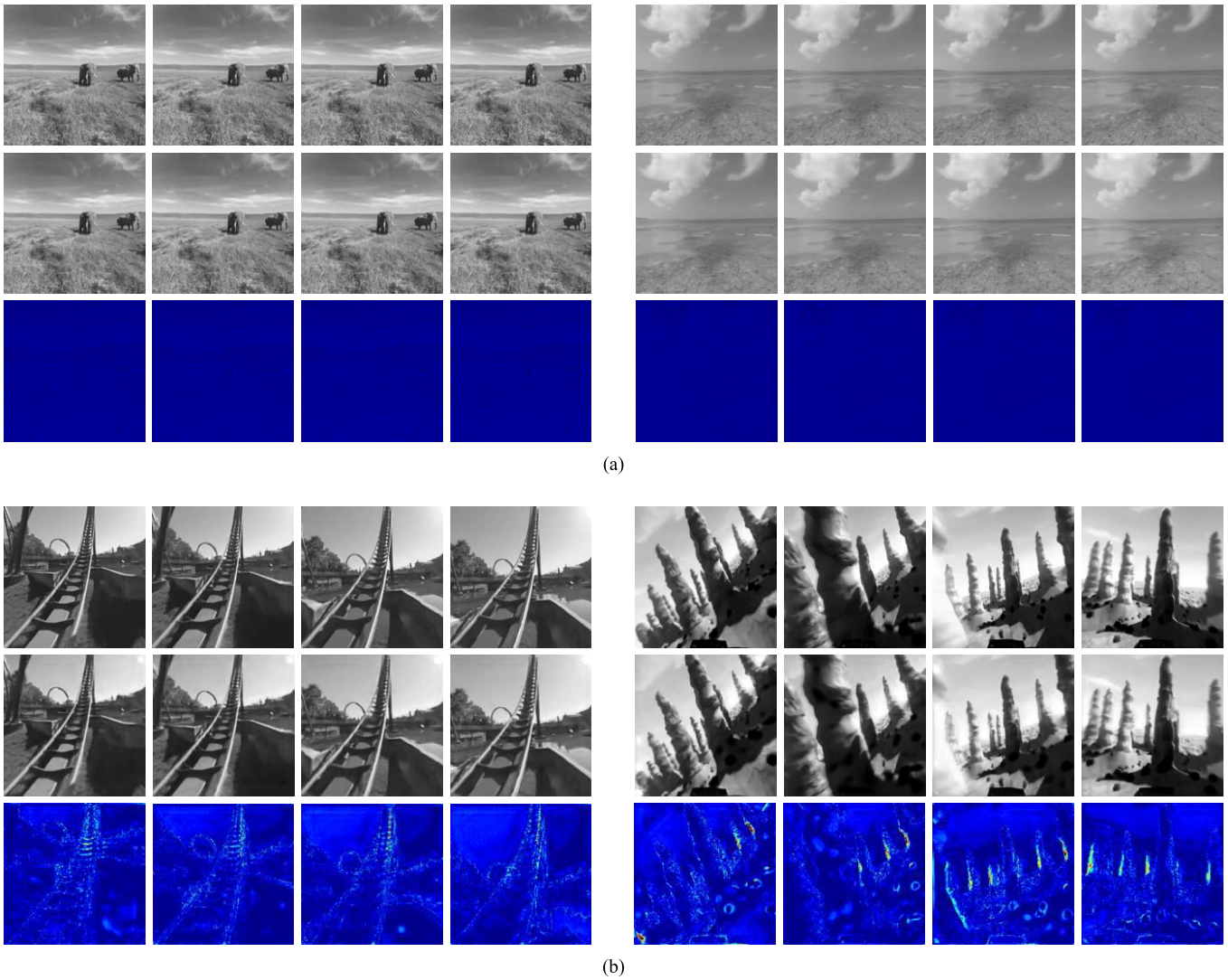


Fig. 7. Visualization of the proposed VRSA method for (a) VR video 1, 2 with slow motion and (b) VR video 7, 9 with exceptional motion (right). First and second rows indicate consecutive original frames and reconstructed frames, respectively. The third row indicates difference maps, $\mathbf{E}_t$, between original frames and reconstructed frames. Note that all images are normalized in range of [0, 1]. Blue indicates '0' and red indicates '1'.

In addition, to quantitatively show the performances of the trained generator for non-exceptional motion videos and exceptional motion videos, we conducted quantitative experiment. For this purpose, we measured the root mean square error (RMSE) between the generated video and the original video for non-exceptional motion and exceptional motion. The average RMSE for non-exceptional videos (Videos $1 \sim 6$ in TABLE III and Videos $1 \sim 16$ in TABLE VIII) is about 3.61 ± 1.70 (mean $\pm$ std). On the other hand, the average RMSE value for exceptional videos (Videos $7 \sim 9$ in TABLE III and Videos $17 \sim 21$ in TABLE VIII) is about 13.70 ± 3.79 (mean $\pm$ std). The results indicate that the generator trained only on non-exceptional motion videos works well for the non-exceptional videos while it does not work well for the exceptional motion videos relatively. In the following subsection, the qualitative results of the generator are shown according to the motion patterns of video.

C. Visual Results

To interpret the performance of the proposed VRSA framework, we visualized the areas of VR video that mostly affected the VR sickness prediction. Figure 7 shows the results of video reconstruction by the proposed generator for Video 1, 2 and Video 7, 9. In Fig. 7 (a) and (b), the first and second rows indicate the original NFOV frames and the reconstructed NFOV frames by the trained generator. The last row represents the reconstruction error maps (i.e., difference maps, $\mathbf{E}_t$). As shown in the Fig. 7(a), Video 1 and 2 with the simple motion pattern were reconstructed well. On the other hand, the generator could not reconstruct Video 7 and 9 well since the trained generator did not encode the exceptional velocity and rotation during training. As shown in Fig. 7(b), it can be recognized that most of the errors occur around the rapidly varying area such as acceleration in Video 7 or rapid turning Video 9. It means that our generator trained by videos with non-exceptional motion can detect the region leading to

TABLE VII

DETAILED DESCRIPTION OF TWENTY 360-DEGREE VIDEOS COLLECTED FROM VIMEO FOR THE TRAINING OF THE GENERATOR

No	Name	Vimeo ID	Resolution	Fps	Time stamp
1	360 VR Wildlife Video, Botswana	235216937	4096×2048	50.02	00:00 - 00:18
2	360° aerial floating	233018732	3840×1920	29.99	00:30 - 01:00
3	360° DRONE VIDEO	207714165	4096×2048	25.01	00:00 - 00:36
4	360VR Lotte Tower Grand Opening Fireworks(South korea)	211468747	4096×2048	29.97	10:40 - 11:10
5	CubeHouse 360 Video	244324478	3840×1920	25.02	00:00 - 00:36
6	DOLE FOOD COMPANY - THE VR EXPERIENCE	208660977	4096×2048	25.02	00:00 - 00:36
7	Driving down the street with the Omni	222764460	4096×2048	24.02	00:00 - 00:38
8	Drone 360	218207046	3840×1920	30.02	00:00 - 00:30
9	Drone 360° - Le canal Nord du Carré Sénart	207769450	4096×2048	25	00:10 - 00:46
10	Fiumanka 2017 360 VR proba	225843046	3840×1920	29.99	00:00 - 00:30
11	HELICOPTER FLIGHT OVER VANCOUVER. Virtual Reality Video 360 VR	209833679	3840×1920	24.01	00:30 - 01:08
12	Mavic on Air 1	254164507	3840×1920	29.97	00:30 - 01:00
13	Panera Bread VR - "Summer"	238820575	4096×2048	29.99	00:00 - 00:30
14	South Africa - a hot air balloon experience in Pilanesberg National Park	212723815	3840×1920	29.97	00:00 - 00:30
15	Spectacular 360° VR flight over the Aletschglacier	208885328	4096×2048	29.97	00:00 - 00:30
16	Study In New Zealand • Kayaking - North Island lakes, kayaking experience	208635041	3840×1920	25.02	00:00 - 00:36
17	Synchrony Financial - Kayak VR	215950486	3840×1920	29.99	00:00 - 00:30
18	Victoria Harbour Ferry Taxi Ride in 4k 360 video for viewing in VR by This Is Me In VR	224271712	3840×1920	29.97	00:10 - 00:40
19	VR360 Himalayan Paragliding 6000m 4K	209944373	3840×1920	30.02	00:10 - 00:40
20	Wild Dolphins VR / 360° Video Experience	208104218	3840×1920	60.02	00:10 - 00:25

Note that the readers can access the dataset using Vimeo ID of each video by entering the Vimeo URL with Vimeo ID: "https://vimeo.com/"+"Vimeo ID".

TABLE VIII

DETAILED DESCRIPTION OF TWENTY ONE 360-DEGREE VIDEOS COLLECTED FROM [59] FOR THE TRAINING OF THE PREDICTOR

Motion type	No	Name	Resolution	Fps	Time stamp
Non-exceptional motion	1	Compilation of various animals and locations, Africa	4096×2048	50.00	00:10 - 01:04
	2	Driving across Golden Gate Bridge, San Francisco, California, USA	4096×2048	29.97	00:00 - 01:30
	3	Gardasee - Strada della Forra	4096×2048	30.00	06:30 - 08:00
	4	Drive to the Rhinos	4096×2048	50.00	00:05 - 00:59
	5	Travelling by car on road, Miami, Florida, USA	4096×2048	29.97	00:00 - 01:30
	6	Safari vehicles on hilltop, overlooking vast landscape, Africa	4096×2048	60.00	00:10 - 00:55
	7	Car travelling along road	4096×2048	50.00	00:00 - 00:54
	8	Male friends driving car towards Gulfoss Waterfall, Iceland, Europe	3840×2160	29.97	00:00 - 01:30
	9	Safari vehicle approaching giraffes, Africa	4096×2048	25.00	00:00 - 01:48
	10	Double-decker bus tour in Valencia, Spain	4096×2048	59.94	00:00 - 00:45
	11	Driving in a BMW, through a City	3840×1920	29.97	00:50 - 02:20
	12	Driving in Manila, Philippines, Asia	3840×1920	29. 96	00:00 - 01:30
	13	Travelling on narrow road by car to Hetch Hetchy, Yosemite Valley, California, USA	4096×2048	29.97	00:00 - 01:30
	14	Off Road Driving in Brazil, South America	3840×1920	29.96	00:00 - 01:30
	15	Journey of car travelling fast along road	4096×2048	50.00	00:00 - 00:54
	16	Car driving through game reserve, South Africa	4096×2048	29.99	00:00 - 01:30
Exceptional motion	17	Motocycle ride	4096×2048	50.00	00:00 - 00:54
	18	Driving car through countryside and roadside buildings	4096×2048	50.00	00:00 - 00:54
	19	Car journey by trees and landscape, USA	4096×2048	50.00	00:00 - 00:54
	20	Time lapse of journey around Melbourne Grand Prix Circuit, Australia	3840×1920	29.97	00:00 - 01:30
	21	Footage of scooter journey through city, Thailand, Asia	3840×1920	29.97	00:10 - 01:40

excessive VR sickness and our predictor can predict the level of VR sickness based on the quality of the generated videos.

VI. DISCUSSIONS

It should be noted that this study was intended for assessing the impact of exceptional motion on VR sickness of 360-degree video for normal vision and healthy people. The literature reported human factors for children and VR sickness-sensitive people [52], [53]. This means that there are limitations in equally applying the proposed method to such subjects with different human factors. Furthermore, it might have to consider human factor in VRSA.

In future work, we will extend the proposed method to assess the VR sickness by considering the human factors (e.g., VR sickness susceptibility). In addition, the other causes of VR sickness need to be further investigated for VR sickness assessment of 360-degree video. It might be helpful for VR sickness assessment in future work.

VII. CONCLUSIONS

In this paper, we proposed a novel objective deep generative model-based VRSA Net for 360-degree videos. In the

proposed method, instead of end-to-end training the regression model with a large number of VR datasets and corresponding subjective scores (i.e., ground truth), the VR video generator based on GAN was devised to learn the tolerance of VR sickness in human motion perception. To encode and decode the characteristics of the VR video with non-exceptional motion, which do not induce excessive VR sickness on human motion perception, we trained the VR video generator only with normal videos with non-exceptional videos. Based on the generated videos by the trained our generator, the proposed VR sickness score predictor could precisely assess the proposed VR sickness score for a test video. In addition, we introduced a benchmark database for the evaluation of VR sickness assessment. We collected nine 360-degree videos including various motion patterns and performed extensive subjective experiments. In our subjective assessment experiment, physiological signals (heart rate and galvanic skin response signals) and subjective questionnaires (SSQ scores) were measured for evaluating VR sickness. In our experiment, the prediction performance showed that the proposed VRSA had a strong correlation with human perception of VR sickness. Furthermore, by visualizing the difference maps between the original and the generated videos by the trained generator, we interpret that the proposed VRSA network quantifies the degree of VR sickness and it could detect area where VR sickness caused by exceptional motion is highly related.

APPENDIX A

For training of the proposed VR video generator, we used twenty 360-degree videos collected from Vimeo. They have slow and constant motion, which could not induce VR sickness. TABLE VII shows the details of twenty 360 videos.

APPENDIX B

For training of the proposed VR sickness score predictor, we used twenty one 360-degree videos collected from [59]. They have various scenes and motion patterns. TABLE VIII shows the details of twenty one 360 videos, which could be found and downloaded from [59] for fee.

REFERENCES

[1] C. Grunheit, A. Smolic, and T. Wiegand, "Efficient representation and interactive streaming of high-resolution panoramic views," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2002, pp. 209–212.

[2] P. R. Alface, J.-F. Macq, and N. Verzijp, "Interactive omnidirectional video delivery: A bandwidth-effective approach," *Bell Labs Tech. J.*, vol. 16, no. 4, pp. 135–147, Mar. 2012.

[3] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness," *Int. J. Aviation Psychol.*, vol. 3, no. 3, pp. 203–220, 1993.

[4] K. Carnegie and T. Rhee, "Reducing visual discomfort with HMDs using dynamic depth of field," *IEEE Comput. Graph. Appl.*, vol. 35, no. 5, pp. 34–41, Sep/Oct. 2015.

[5] 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*, vol. 29, no. 2, pp. 58–69, Mar. 2008.

[6] K. W. Arthur, "Effects of field of view on performance with head-mounted displays," Ph.D. dissertations, Dept. Comput. Sci., Univ. North Carolina Press, Chapel Hill, NC, USA, 2000.

[7] J. J.-W. Lin, H. B. L. Duh, D. E. Parker, H. Abi-Rached, and T. A. Furness, "Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment," in *Proc. IEEE Virtual Reality*, Orlando, FL, USA, Mar. 2002, pp. 164–171.

[8] A. S. Fernandes and S. K. Feiner, "Combating VR sickness through subtle dynamic field-of-view modification," in *Proc. IEEE Symp. 3D User Interfaces*, Greenville, SC, USA, Mar. 2016, pp. 201–210.

[9] R. S. Allison, L. R. Harris, M. Jenkin, U. Jasiobedzka, and J. E. Zacher, "Tolerance of temporal delay in virtual environments," in *Proc. IEEE Virtual Reality (VR)*, Yokohama, Japan, Mar. 2001, pp. 247–254.

[10] S. Kawamura and R. Kijima, "Effect of head mounted display latency on human stability during quiescent standing on one foot," in *Proc. IEEE Virtual Reality (VR)*, Greenville, SC, USA, Mar. 2016, pp. 199–200.

[11] H. B. L. Duh, D. E. Parker, J. O. Philips, and T. A. Furness, "'Conflicting' motion cues to the visual and vestibular self- motion systems around 0.06 Hz evoke simulator sickness," *Human Factors*, vol. 46, no. 1, pp. 142–153, 2004.

[12] L. J. Hettinger and G. E. Riccio, "Visually induced motion sickness in virtual environments," *Presence, Teleoper. Virtual Environ.*, vol. 1, no. 3, pp. 306–310, 1992.

[13] E. L. Groen and J. E. Bos, "Simulator sickness depends on frequency of the simulator motion mismatch: An observation," *Presence*, vol. 17, no. 6, pp. 584–593, Dec. 2008.

[14] S. V. Mammen, A. Knot, and S. Edenhofer, "Cyber sick but still having fun," in *Proc. 22nd ACM Conf. Virtual Reality Softw. Technol. (VRST)*, New York, NY, USA, Nov. 2016, pp. 325–326.

[15] B. Keshavarz, B. E. Riecke, L. J. Hettinger, and J. L. Campos, "Vection and visually induced motion sickness: How are they related?" *Frontiers Psychol.*, vol. 6, no. 472, p. 472, 2015.

[16] A. M. Gavgani, D. M. Hodgson, and E. Nalivaiko, "Effects of visual flow direction on signs and symptoms of cybersickness," *PLoS ONE*, vol. 12, no. 8, pp. 1–4, Aug. 2017.

[17] A. Pavel, B. Hartmann, and M. Agrawala, "Shot orientation controls for interactive cinematography with 360 degree video," in *Proc. 30th Annu. ACM Symp. User Interface Softw. Technol. (UIST)*, Aug. 2017, pp. 289–297.

[18] Y. Y. Kim, H. J. Kim, E. N. Kim, H. D. Ko, and H. T. Kim, "Characteristic changes in the physiological components of cybersickness," *Psychophysiology*, vol. 42, no. 5, pp. 616–625, Sep. 2005.

[19] M. S. Dennison, A. Z. Wisti, and M. D. Zmura, "Use of physiological signals to predict cybersickness," *Displays*, vol. 44, pp. 42–52, Sep. 2016.

[20] D. Egan, S. Brennan, J. Barrett, Y. Qiao, C. Timmerer, and N. Murray, "An evaluation of heart rate and electrodermalactivity as an objective QoE evaluation method for immersive virtual reality environments," in *Proc. 8th Int. Conf. Qual. Multimedia Exper. (QoMEX)*, Lisbon, Portugal, 2016, pp. 1–6.

[21] M. Meehan, B. Insko, M. Whitton, and F. P. Brooks, "Physiological measures of presence in stressful virtual environments," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 645–652, Jul. 2002.

[22] J.-P. Stauffert, F. Niebling, and M. E. Latoschik, "Towards comparable evaluation methods and measures for timing behavior of virtual reality systems," in *Proc. 22nd ACM Conf. Virtual Reality Softw. Technol. (VRST)*, New York, NY, USA, Nov. 2016, pp. 47–50.

[23] M. Chessa, G. Maiello, A. Borsari, and P. J. Bex, "The perceptual quality of the oculus rift for immersive virtual reality," *Hum.-Comput. Interact.*, pp. 1–32, Dec. 2016.

[24] S. Palmisano, R. Mursic, and J. Kim, "Vection and cybersickness generated by head-and-display motion in the Oculus rift," *Displays*, vol. 46, pp. 1–8, Jan. 2017.

[25] *KAIST IVY Lab. Database*. Accessed: 2018. [Online]. Available: https://ivylabdb.kaist.ac.kr

[26] M. Yu, H. Lakshman, and B. Girod, "A framework to evaluate omnidirectional video coding schemes," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Sep./Oct. 2015, pp. 31–36.

[27] Y. Sun, A. Lu, and L. Yu, "Weighted-to-spherically-uniform quality evaluation for omnidirectional video," *IEEE Sig. Proces. Lett.*, vol. 24, no. 9, pp. 1408–1412, Sep. 2017.

[28] V. Zakharchenko, K. P. Choi, and J. H. Park, "Quality metric for spherical panoramic video," *Proc. SPIE*, vol. 9970, p. 99700C, Sep. 2016.

[29] H.-T. Lim, H. G. Kim, and Y. M. Ro, "VR IQA NET: Deep virtual reality image quality assessment using adversarial learning," in

Proc.IEEE Int. Conf. Acoust. Speech Sig. Process. (ICASSP), Apr. 2018, pp. 6737–6741.

[30] A. Singla, S. Fremerey, W. Robitza, and A. Raake, "Measuring and comparing QoE and simulator sickness of omnidirectional videos in different head mounted displays," in *Proc. 9th Int. Conf. Qual. Multimedia Exper. (QoMEX)*, Erfurt, Germany, May/Jun. 2017, pp. 1–6.

[31] Y. J. Jung, H. Sohn, S.-I. Lee, H. W. Park, and Y. M. Ro, "Predicting visual discomfort of stereoscopic images using human attention model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 12, pp. 2077–2082, Dec. 2013.

[32] J. Park, H. Oh, S. Lee, and A. C. Bovik, "3D visual discomfort predictor: Analysis of disparity and neural activity statistics," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1101–1114, Mar. 2015.

[33] H. Sohn, Y. J. Jung, S.-I. Lee, and Y. M. Ro, "Predicting visual discomfort using object size and disparity information in stereoscopic images," *IEEE Trans. Broadcast.*, vol. 59, no. 1, pp. 28–37, Mar. 2013.

[34] Y. J. Jung, H. G. Kim, and Y. M. Ro, "Critical binocular asymmetry measure for the perceptual quality assessment of synthesized stereo 3D images in view synthesis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 7, pp. 1201–1214, Jul. 2016.

[35] F. Shao, W. Lin, S. Gu, G. Jiang, and T. Srikanthan, "Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1940–1953, May 2013.

[36] J. Kim and S. Lee, "Deep learning of human visual sensitivity in image quality assessment framework," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, Jul. 2017, pp. 1969–1977.

[37] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, and W. Zuo, "End-to-end blind image quality assessment using deep neural networks," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1202–1213, Mar. 2018.

[38] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 206–219, Jan. 2018.

[39] H. Oh, S. Ahn, J. Kim, and S. Lee, "Blind deep S3D image quality evaluation via local to global feature aggregation," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4923–4936, Oct. 2017.

[40] H. Jeong, H. G. Kim, and Y. M. Ro, "Visual comfort assessment of stereoscopic images using deep visual and disparity features based on human attention," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 715–719.

[41] H. G. Kim, H. Jeong, H. Lim, and Y. M. Ro, "Binocular fusion net: Deep learning visual comfort assessment for stereoscopic 3D," *IEEE Trans. Circuits Syst. Video Technol.*, to be published.

[42] J. T. Reason, "Motion sickness adaptation: A neural mismatch model," *J. Roy. Soc. Med.*, vol. 71, no. 11, pp. 819–829, Nov. 1978.

[43] J. E. Bos, W. Bles, and E. L. Groen, "A theory on visually induced motion sickness," *Displays*, vol. 29, no. 2, pp. 47–57, Mar. 2008.

[44] I. M. Arafat, S. M. S. Ferdous, and J. Quarles, "The effects of cybersickness on persons with multiple sclerosis," in *Proc. 22nd ACM Conf. Virtual Reality Softw. Technol. (VRST)*, New York, NY, USA, Nov. 2016, pp. 51–59.

[45] Y-C. Su, D. Jayaraman, and K. Grauman, "Pano2Vid: Automatic cinematography for watching 360 degree videos," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, Nov. 2016, pp. 154–171.

[46] K. Simonyan and A. Zisserman. (Sep. 2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556

[47] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2015, pp. 1520–1528.

[48] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2015, pp. 802–810.

[49] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2014, pp. 2672–2680.

[50] Y. Rai and P. L. Callet, "A dataset of head and eye movements for 360 degree images," in *Proc. ACM Multimedia Syst. Conf.*, 2017, pp. 205–210.

[51] X. Corbillon, F. D. Simone, and G. Simon, "360-degree video head movement dataset," in *Proc. ACM Multimedia Syst. Conf.*, Jun. 2017, pp. 199–204.

[52] J. Hakkinen, T. Vuori, and M. Paakka, "Postural stability and sickness symptoms after HMD use," in *Proc. IEEE Int. Conf. Syst. Man Cybern.*, Oct. 2002, pp. 147–152.

[53] G. D. Park, R. W. Allen, D. Fiorentino, T. J. Rosenthal, and M. L. Cook, "Simulator sickness scores according to symptom susceptibility, age, and gender for an older driver assessment study," in *Proc. Human Factors Ergonom. Soc. Annu. Meeting*, Oct. 2006, pp. 2702–2706.

[54] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document ITU-R BT.500-13, 2012.

[55] *Subjective Methods for the Assessment of Stereoscopic 3DTV Systems*, document ITU-R BT.2021, 2012.

[56] S. Bruck and P. A. Watters, "Estimating cybersickness of simulated motion using the simulator sickness questionnaire SSQ: A controlled study," in *Proc. 6th Int. Conf. Comput. Graph. Imag. Vis.*, Tianjin, China, Aug. 2009, pp. 486–488.

[57] B. Keshavarz and H. Hecht, "Validating an efficient method to quantify motion sickness," *Human Factors*, vol. 53, no. 4, pp. 415–426, Aug. 2011.

[58] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, Sep. 2013.

[59] *Blend Media The Global 360 Video*. Accessed: 2018. [Online]. Available: https://blend.media/

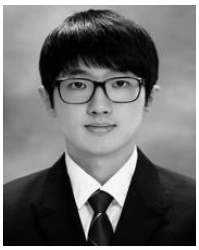
[60] D. P. Kingma and L. J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.(ICLR)*, 2015, pp. 1–13.

[61] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[62] *Methods, Metrics and Procedures for Statistical Evaluation, Qualification and Comparison of Objective Quality Prediction Models*, document ITU-T p.1401, 2012.

Hak Gu Kim received the B.S. and M.S. degrees from Inha University, Incheon, South Korea, in 2012 and 2014, respectively. He is currently pursuing the Ph.D. degree with the Korea Advanced Institute of Science and Technology, Daejeon, South Korea. His research interests include deep learning, virtual reality (VR), 3D image/video processing, human 3D/VR perception, and visual quality assessment.

Heoun-Taek Lim received the B.S. degree from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2015, where he is currently pursuing the M.S. degree. His research interests include deep learning, virtual reality (VR), 3D image/video processing, human 3D/VR perception, and visual quality assessment.



Sangmin Lee received the B.S. degree from Yonsei University, Seoul, in 2017. He is currently pursuing the Joint M.S./Ph.D. degree with the Korea Advanced Institute of Science and Technology, Daejeon, South Korea. His research interests include deep learning, image and video analysis, and visual quality assessment.

Yong Man Ro (S'85–M'92–SM'98) received the B.S. degree from Yonsei University, Seoul, South Korea, and the M.S. and Ph.D. degrees from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea. He was a Researcher with Columbia University, a Visiting Researcher with the University of California, Irvine, CA, USA, and a Research Fellow with the University of California, Berkeley, CA, USA. He was a Visiting Professor with the Department of Electrical and Computer Engineering, University of Toronto, Canada. He is currently a Professor with the Department of Electrical Engineering, KAIST. He established the Image and video Systems (IVY) Lab, KAIST, in 1997. Among the years, he has been conducting research in a wide spectrum of image and video systems research topics. His recent research interests are deep learning, machine learning in computer vision and image processing (2D, 3D, VR), medical imaging, visual recognition, and visual quality assessment. He was a recipient of the Young Investigator Finalist Award of ISMRM in 1992 and the Year's Scientist Award, South Korea, in 2003. He served as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS. He serves as an Associate Editor for the *Transactions on Data Hiding and Multimedia Security* (Springer-Verlag). He served for TPC in many international conferences, including the Program Chair, and organized special sessions.