VR Sickness Versus VR Presence: A Statistical Prediction Model

Woojae Kim, Sanghoon Lee, *Senior Member, IEEE*, and Alan Conrad Bovik, *Fellow, IEEE*

***Abstract*— Although it is well-known that the negative effects of VR sickness, and the desirable sense of presence are important determinants of a user's immersive VR experience, there remains a lack of definitive research outcomes to enable the creation of methods to predict and/or optimize the trade-offs between them. Most VR sickness assessment (VRSA) and VR presence assessment (VRPA) studies reported to date have utilized simple image patterns as probes, hence their results are difficult to apply to the highly diverse contents encountered in general, real-world VR environments. To help fill this void, we have constructed a large, dedicated VR sickness/presence (VR-SP) database, which contains 100 VR videos with associated human subjective ratings. Using this new resource, we developed a statistical model of spatio-temporal and rotational frame difference maps to predict VR sickness. We also designed an exceptional motion feature, which is expressed as the correlation between an instantaneous change feature and averaged temporal features. By adding additional features (visual activity, content features) to capture the sense of presence, we use the new data resource to explore the relationship between VRSA and VRPA. We also show the aggregate VR-SP model is able to predict VR sickness with an accuracy of 90% and VR presence with an accuracy of 75% using the new VR-SP dataset.**

***Index Terms*— VR sickness assessment (VRSA), VR presence assessment (VRPA), natural video statistics (NVS), human visual system (HVS).**

I. INTRODUCTION

In RECENT decades, there have been tremendous advances in the development of virtual reality (VR) technologies [1]. VR devices have been successfully deployed as a way of simulating real-world experiences in a wide variety of domains including gaming, simulators and medical clinics, using increasingly lightweight, comfortable, and immersive head-mounted displays (HMD). However, the quality of experience (QoE) of VR users is often severely reduced by VR sickness. In many ways, providing a satisfactory and realistic sense of presence to VRs users involves navigating trade-offs between immersion and comfort.

Here we focus on the development of ways to navigate these trade-offs. Specifically, we develop a model of VR sickness assessment (VRSA) and also develop a VR presence assessment (VRPA) model that takes advantage of predicted VR sickness scores, inspired by the trade-off relationship between VR sickness and VR presence. This allows the model to predict an optimal combination of factors contributing to satisfaction in VR.

In many previous studies, VR sickness has been reported to induce oculomotor symptoms such as visual fatigue and difficulty focusing; physiological reactions including burping, salivation and sweating; and disorientation, dizziness, and vertigo [2]–[4]. One of the most common causes of VR sickness is a sensory mismatch between the vestibular system and the visual system. The vestibular system, which perceives actual movement, and the visual system, which perceives projected motion fields, may provide conflicting visual percepts [5], [6]. When there is a visual perception of movement, especially self-motion, that is experienced in a static physical state, visual-vestibular conflicts may arise in the brain. Because of this, a number of authors have reported ways of measuring VR sickness by quantifying the amount of perceptual motion. For example, the authors of [7] predict VR sickness by calculating differences between perceived motions and head movements using a visual-vestibular conflict model. In [8], a deep autoencoder based model is developed that predicts exceptional motion. The model is also generalized using a generative adversarial network (GAN) [9].

However, these studies have not benefited from the availability of sizeable labeled datasets. We utilize a new, large database that we have created to develop a new, highly competitive model that is capable of predicting both the degree of VR sickness that may be felt, as well as the VR sense of presence. This dual model may prove useful for mediating these negative and positive aspects during content creation or display.

Our approach to VRSA differs from prior methods in that it seeks to quantify losses of statistical regularity of VR content that are predictive of sickness arising from VR content. Specifically, the contributions that we make are summarized:

1) We built the first large database addressing both VR sickness and VR presence, including subjective labels.
2) We develop new predictive models for VRSA and VRPA that are based on natural VR content statistics.
3) We demonstrate state-of-the-art performance of the new models for VR sickness and VR presence prediction.

Manuscript received April 7, 2020; revised August 28, 2020; accepted October 19, 2020. Date of publication November 18, 2020; date of current version November 24, 2020. This work was supported by the Samsung Research Funding Center of Samsung Electronics under Project SRFC-IT1702-08. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Joao M. Ascenso. *(Corresponding author: Sanghoon Lee.)*

Woojae Kim and Sanghoon Lee are with the Department of Electrical and Electronic Engineering, Yonsei University, Seoul 120-749, South Korea (e-mail: wooyoa@yonsei.ac.kr; slee@yonsei.ac.kr).

Alan Conrad Bovik is with the Laboratory for Image and Video Engineering, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084 USA (e-mail: bovik@ece.utexas.edu).

Digital Object Identifier 10.1109/TIP.2020.3036782

The remainder of the paper is organized as follows. Section II introduces related works on VR sickness and presence prediction. Section III describes the design of the proposed VR-SP database. Section IV introduces the VRSA based on predictive statistical features, while Section VI describes our VRPA model, which uses the predicted VR sickness scores, along with measures of visual activity and content. Section V discusses experimental results, and conclusions are given in Section VII.

II. RELATED WORKS

A. Sickness Assessment in VR Environment

There have been several studies of VR sickness from a physiological or behavioral perspective [10]–[13]. Most of these have involved simulator sickness questionnaires (SSQ) or physiological tests such as electroencephalogram (EEG), galvanic skin response (GSR), electrogastrogram (EGG), or heart rate [10]–[12]. However, these sensor modalities are quite noisy, and it is difficult to engage subjects in viewing and rating a large amount of VR contents under these protocols. Moreover, these tests have typically deployed simple visual patterns (e.g., stripes) [6], [14], [15].

However, content that is being viewed on modern VR devices, such as the Oculus Rift/Quest and HTC-VIVE HMDs, is trending towards increased realism and naturalness, along with more immersive experiences. Being able to automatically and objectively predict VR sickness in HMD environments has become highly desirable.

B. Natural Scene Statistics

A number of natural scene statistics (NSS) methods have been developed to measure losses of statistical regularities arising from distortions of picture and videos [16]. The efficacy of this approach has been amply demonstrated by the development of highly distortion-sensitive bandpass decomposition methods in the wavelet domain [17], the discrete cosine transform (DCT) [18], and in spatial [19] and frame difference [20] domains. Inspired by these advances, we have developed methods for predicting VR sickness from variations in immersive video statistics.

C. Sense of Presence

There has been a variety of studies of the sense of presence experienced when viewing 2D and stereoscopic 3D (S3D) visual content. Early on, Ijsselsteijn *et al.,* introduced the concept of visual presence and how to measure it [21]. They also conducted a study of subjective presence on S3D content, and derived objective measures of the sense of presence [22]. They also gathered subjective opinions of both presence and discomfort of viewers of stereoscopic cinema. Oh and Lee [23] developed a visual presence assessment model utilizing the geometry of S3D displays.

In connection with VR sickness, it is generally thought that VR sickness is one of the most crucial factors causing users to feel differences between virtuality and reality [24]–[28]. When VR sickness in virtual space exceeds that experienced in the real world, then the sense of realism is reduced [29]. However, the relationship between VR sickness and VR presence has not yet been quantified adequately, because of the lack of a sufficiently large VR content database with subject labels of presence and sickness. Because of this, the problem of predicting VR presence and VR sickness, and the relationship between them, remains unresolved.

III. VR-SP DATABASE

In this section, we describe the new VR-SP database. It is well known that both VR sickness and presence are regarded as critical factors that affect the popularity and viability of VR products. Existing databases of psychometric scores are quite limited in the number of included VR contents [8], [9], [30]. Moreover, most of these databases focus only on VR sickness, with little or no analysis of VR presence or its relationship with VR sickness. Towards advancing progress in this direction, our new VR-SP database contains a wide variety of diverse VR source contents, along with corresponding subjective scores on VR sickness and VR presence. The VR-specific content in the database can be viewed on any popular VR device, such as the HTC-VIVE or the Oculus Rift.

A. VR Content

The new VR content database contains 10 reference VR scenes providing a variety of virtual experiences (*e.g.,* VR games, roller-coaster, and outer space), as tabulated in Table I. To create diverse experiences of VR sickness, highly diverse camera rotations (yaw, pitch, and roll) and translations (forward, backward, and lateral) were deployed, depending on each scene. All the reference contents were implemented with various presets on the Unity platform.

Since our goal is to design a database enabling the analysis and quantification of VR sickness and VR presence, a total of 100 VR contents were created from the 10 reference VR scenes. Toward this, each reference content was diversified to 10 variations by combining two types of motion, and four levels of velocity, applied to the 10 reference VR contents in Fig. 1. In addition to spatially diversifying the VR contents, we reduced the level of detail in the scenes. The detail-reduced scenes were those containing only the two lower levels of velocity. Fig. 2 diagrams the variations (V1-V10) of each reference scene. As shown in the figure, the motion types and velocity levels were applied to the individual reference and detail-reduced contents. The reference videos have detailed environments, as shown in Fig. 3, while the modified scenes contain less detail. The VR-SP database also contains two types of motion. The first motion category is a simple movement, such as linear motion, while the second includes complex motions, such as rotations and rapid transitions. Each motion type includes four levels of velocity, while for the detail-modified scenes, only motion type 1 was used at velocity levels 1 and 2.

After processing them with the various motion types and velocity levels mentioned above, a subjective study was conducted, wherein all of the VR sequences were viewed

TABLE I

VR CONTENT IN THE VR-SP DATABASE

Scenario #	Name	Context
#1	Horse	Riding a horse on the forest.
#2	Centrifuge	Experiencing centrifuge in the dome.
#3	Boat	Riding a boat in the rapids.
#4	Drift	Experiencing rapid by relying on driftwood.
#5	AirFighter	Riding an air fighter in the cyberspace.
#6	BounceBall	Experiencing bouncing ball in the cyberspace.
#7	Car	Riding a car on the outdoor space.
#8	RollerCoaster	Riding a rollercoaster on the outdoor space.
#9	Indoor	Walking in the indoor space.
#10	Space	Experiencing spacewalk near the space station.

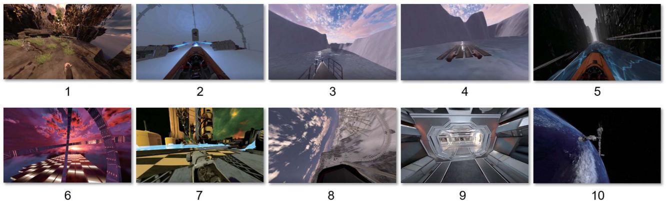


Fig. 1. Exemplar reference scenes in the VR-SP database.

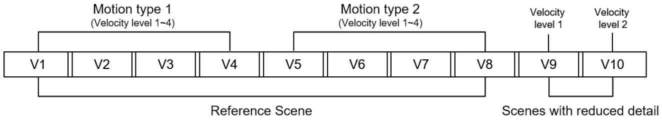
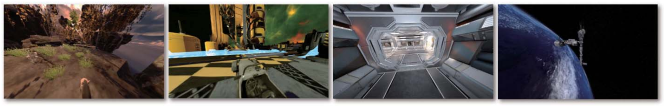
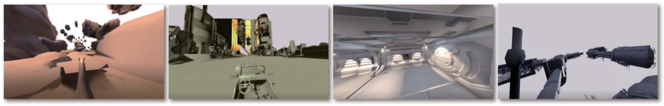


Fig. 2. 10 variations (V1-V10) of each scenario with two motion types and four velocity levels on the reference scene and scenes with reduced detail, in the VR-SP database.



Reference Scenes



Scenes with reduced detail

Fig. 3. Exemplars of reference and their reduced detail scenes in the VR-SP database.

by 21 human subjects wearing an HMD device. Each 60 fps VR video sequence has a resolution of 1080 × 600, 24-bits/pixel color frames and a duration of 9 seconds. All of the sequences were encoded by the baseline H.264 Unity video coder.

B. Subjective Assessment

As recommended in the ITU-R BT.500-13 [31] and BT.2021 [32] standards, we adopt a single-stimulus subjective scoring methodology for both sickness and presence assessments. The 21 inexperienced subjects (satisfying the subject criteria recommended in [31]) were of ages ranging from 22 to 34 years. All of the subjects were screened for normal visual acuity on the Landolt chart. The overall protocol included four evaluation sessions, each containing 25 randomly shuffled VR contents from the database. After each individual session, rest periods of 5 min duration were inserted, to minimize any accumulated feeling of VR sickness [33]. In each session, a VR sequence was displayed for 9 seconds, after which the subject assessed their feelings of sickness and presence. Similar to previous work [34], VR presence is defined here as the subjective sense of presence ("actually being there") in the user's space or environment. To measure VR sickness, a simple scale of experienced sickness was used rather than using a lengthy questionnaire, which would have subjected the subjects to unnecessarily long sessions (several hours), and would have made it impossible to collect large amounts of data. This was facilitated by a user controller-based VR interface that allowed the subjects to interactively score in the same display environment. The subjects rated the contents on a discrete, 5-points Likert scale marked as follows.

We and others [35] have found that viewers do experience VR sickness on such short time scales, and indeed early sensations of sickness may be the most important to detect, before they accumulate and become more severe.

The VR Sickness labels on the scale were: Extremely Uncomfortable (5), Uncomfortable (4), Mildly Comfortable (3), Comfortable (2), and Very Comfortable (1). The VR Presence labels were Excellent (5), Good (4), Fair (3), Poor (2), and Bad (1).

The parenthetical numbers indicate the numerical associations with the labels. After all the subjects scored the videos, the mean opinion scores (MOS) of both sickness/presence were obtained. Fig. 4 plots the distribution of MOS values of both VR sickness and VR presence. As shown in the figure, the distribution is not biased to a specific score for each target value. We summarize major information about the test environment in Table II.

C. Trade-Offs Between VR Sickness and VR Presence

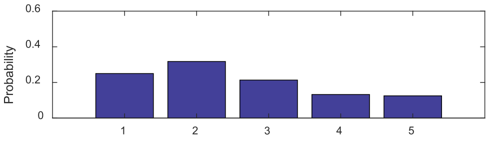
Fig. 5 shows a scatter diagram and fitted curve of the MOS scores of VR sickness and VR presence plotted against each other, over all samples on the VR-SP database. Interestingly, the VR presence score is gradually lowered in the very low VR sickness region (MOS < 0.4). Furthermore, in the region with excessively high VR sickness (MOS > 0.8), the VR presence scores also trended gradually lower. However, most of the areas with high VR presence were accompanied by moderate VR sickness scores. This relationship may be more clearly seen by the fitted curve. By observation, the most intuitive way to predict VR presence is to discover a specific range where the user experiences moderate VR sickness and proper immersion. Therefore, we will use the predicted VR sickness score as a feature of VRPA in Section V.

IV. PROPOSED VR SICKNESS ASSESSOR

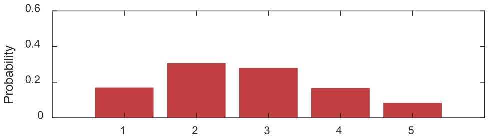
Most VRSA studies have focused on understanding VR sickness, by using motion estimation to probe the level of visual-vestibular sensory conflict. Here we instead use natural video statistic (NVS) models derived in the context of a set

TABLE II
VR-SP DATABASE CONFIGURATION AND SUBJECTIVE EXPERIMENTAL CONDITIONS

Simulation platform	Unity
Color depth	24-bits/pixel color frames
Video coder	MPEG-4
Subjects	21 inexperienced subjects
Frame resolution	1080×600
Duration	9 seconds
# of video	100
Viewing environment	HTC-VIVE
Method	Single-stimulus (SS)
Evaluation scales	Likert-scale (1 to 5)



(a) Sickness MOS



(b) Presence MOS

Fig. 4. Histograms of MOS values for (a) VR sickness and (b) VR presence in the VR-SP database.

of perceptually relevant processes. Moreover, we devise an exceptional motion feature using a measurement of the correlation between statistical features drawn from instantaneous and averaged temporal processing. The design of our VR sickness model is inspired by prior NVS approaches [20], [36] whereby the losses of statistical regularities arising from undesirable visual content, such as distortions. The overall flow of the VRSA framework is depicted in Fig. 6.

A. Preprocessing

To realize the most vivid VR experiences, providing a wide field of view (FoV) is an important factor. Most HMD devices are designed with rectangular displays via a lens system, hence barrel distortion is unavoidable, leading to geometric distortions along the image border.

For example, Fig 7 shows an example of a rendered VR content, and its distortion corrected version. As shown in Fig. 7 (a) (red boxes), borderline pixels are stretched compared to those in the central regions. Accordingly, when a motion estimation method or an optical flow algorithm is applied, prediction errors arising from lens distortion occur. To overcome this, we geometrically correct each rendered frame via a pincushion transformation, as shown in Fig. 7 (b) (See red boxes in Fig. 7 (b)) [37]. This also matches the corrected image that viewer sees.

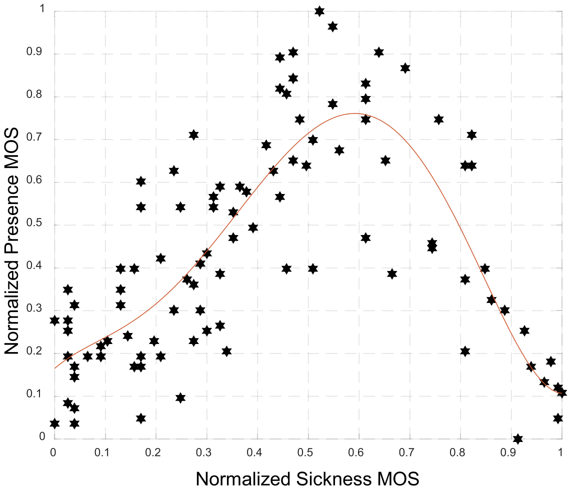


Fig. 5. Scatter plot of normalized subjective scores of VR sickness and VR presence (along with a curve fit) from the VR-SP database. All MOS values were scaled to [0, 1], where 1 equates to stronger feelings of sickness or presence.

B. Spatio-Temporal Frame Differences

After preprocessing the VR sequence, we have luminance frames $\{F^1, F^2, \ldots, F^T\}$ of dimensions $\mathbb{N}^{H\times W\times T}$, where H, W and T are the height, width and total number of frames, respectively. Given each VR video, we do not explicitly estimate motion, but instead compute temporal changes of luminance. As discussed in [38] temporal frame differences obey regular statistical laws, while optical flow/motion vectors are much less regular. The time-averaged p-norm frame difference is defined as:

$$\Delta F_p^t(\mathbf{x}) = |F^\tau(\mathbf{x}) - F^{\tau+1}(\mathbf{x})|^p \tag{1}$$

where $\mathbf{x}$ represents a tuple of spatial indices ($\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2\}$, $\mathbf{x}_1 \in [1, H]$ and $\mathbf{x}_2 \in [1, W]$), over a set of consecutive frames indexed t.

C. Rotational Frame Differences

From previous studies, it is well known that rotational image motions, such as roll or pitch, are highly related to VR sickness [39], [40]. One feasible strategy to model rotational statistics is to reconstruct rotational representations using the frame difference maps, then estimate statistical features on them. Fig. 8 shows nine directional modes with corresponding angles and their reconstructed rotational motion maps (*i.e.,* roll motion, zoom-in/out, horizontal and vertical motions). Fig. 8 (a), we first calculate spatially displaced frame differences over frames, depending on the directional mode. As shown in [41], spatially-displaced frame differences are an effective way to capture space-time statistical regularities predictive of distortion. At each frame index t and spatial index $\mathbf{x}$, spatially displaced frame difference maps are computed as

$$\Delta F_p^t(\mathbf{x}; \theta) = |F^\tau(\mathbf{x}) - F^{\tau+1}(\mathbf{w}(\mathbf{x}; \mathbf{a}_\theta))|^p \tag{2}$$

where $\mathbf{w}(\cdot; \mathbf{a}_\theta)$ is a warping function over the vector of mode parameters dictated by the angle θ. We use eight mode directions to represent motion elements, thus

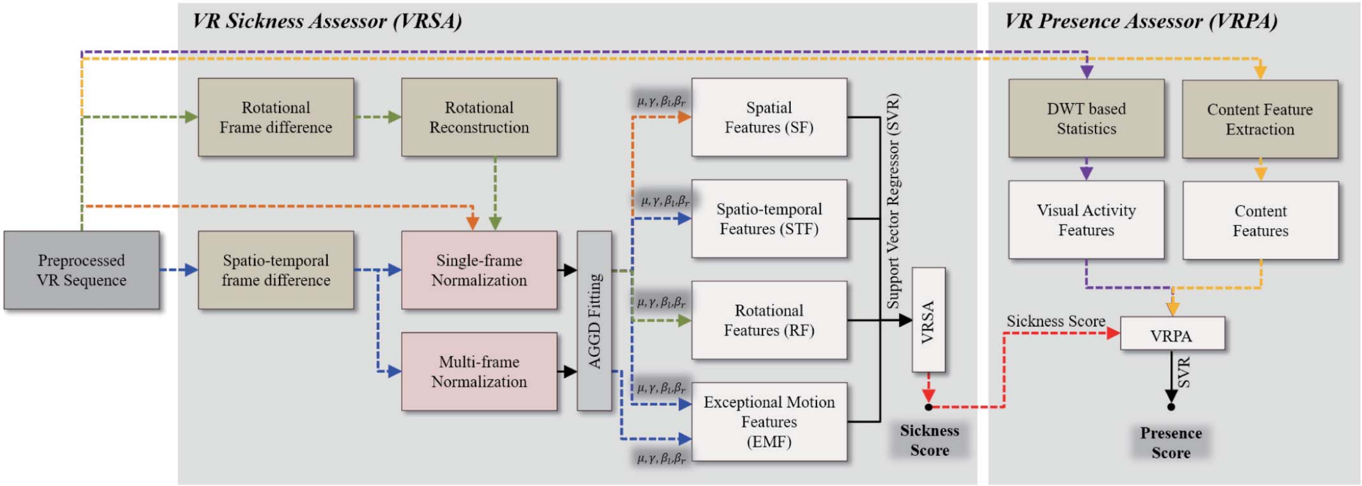


Fig. 6. Block diagram of the proposed sickness and presence predictor.

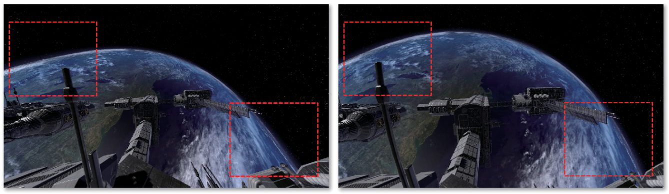

(a) (b)

Fig. 7. Example of geometric preprocessing. (a) original "Space" scene in the VR-SP database; (b) geometric distortion corrected version.

the parameters are obtained as: $(\mathbf{a}_{0^\circ}, \mathbf{a}_{45^\circ}, \ldots, \mathbf{a}_{315^\circ}) = ([1, 0], [1, 1], \ldots, [1, -1])$. Let ΔF denote the spatio-temporal frame difference map, obtained by the same procedure in (1). Then the eight spatially displaced frame difference maps with their corresponding modes are $\{H_1, D_1, V_1, D_2, H_2, D_3, V_2, D_4\}$.

To represent a rotational frame difference, first divide the difference map of each mode into nine local patches. Then, as shown in Figs. 8 (b)-(e), the corresponding local patch of each mode is reconstructed to capture roll, zoom-in/out, horizontal and vertical motions (Figs. 8 (b)$\rightarrow(R_1, R_2)$, (c)$\rightarrow(Z_1, Z_2)$, (d)$\rightarrow(H_1, H_2)$ and (e)$\rightarrow(V_1, V_2)$). Each rotational motion includes directional coefficient pairs, and the final rotational frame difference is defined as the sum of each pair: $\Delta R = R_1 + R_2$, $\Delta Z = Z_1 + Z_2$, $\Delta H = H_1 + H_2$, $\Delta V = V_1 + V_2$.

D. Natural Video Statistics

1) Frame Difference Normalization: Given a frame differenced sequence ΔF^t, the mean subtracted contrast normalization (MSCN) coefficients of it are obtained as [42]:

$$\Delta \hat{F}^t(\mathbf{x}) = \frac{\Delta F^t(\mathbf{x}) - \Delta \mu^t(\mathbf{x})}{\Delta \sigma^t(\mathbf{x}) + C}, \tag{3}$$

over spatial indices ($\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2\}$ where $\mathbf{x}_1 \in [1, H]$ and $\mathbf{x}_2 \in [1, W]$), and over a set of consecutive frame time samples (frame indices), $t \in [1, T]$ where

$$\Delta \mu^t(\mathbf{x}) = \sum_{k=-K}^{K} \sum_{l=-L}^{L} w_{k,l} \Delta F^t(\mathbf{x}_1 + k, \mathbf{x}_2 + l) \tag{4}$$

and

$$\Delta \sigma^t(\mathbf{x}) = \sqrt{\sum_{k=-K}^{K} \sum_{l=-L}^{L} w_{k,l}[\Delta F^t(\mathbf{x}_1 + k, \mathbf{x}_2 + l) - \Delta \mu(\mathbf{x})]^2} \tag{5}$$

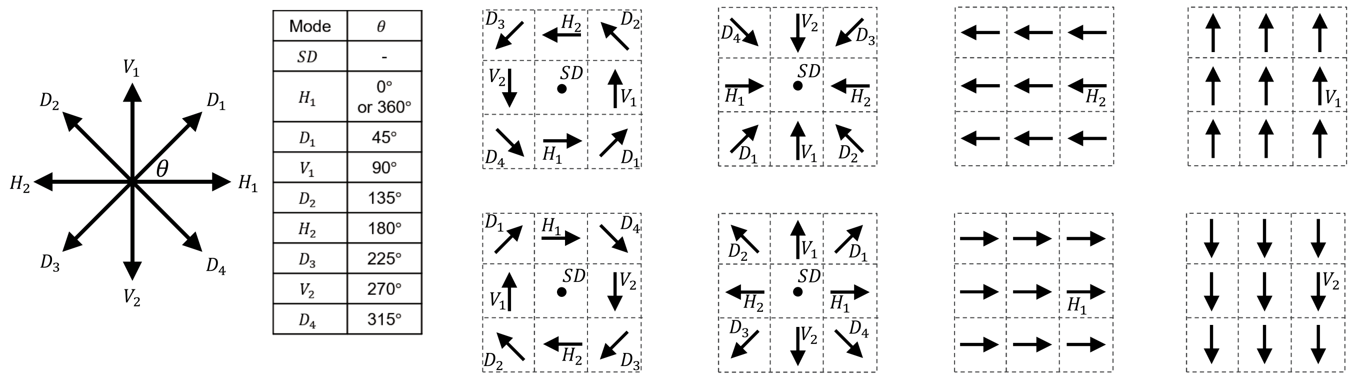
denote the weighted local mean and weighted contrast of each frame difference map, respectively, where $w_{k,l}$ is a Gaussian weighting function sampled out to 3 standard deviations and rescaled to unit volume over ($k = -K, \ldots, K$), ($l = -L, \ldots, L$). In our experiments, we fixed the semi-saturation constant in the divisive normalization to $C = 0.01$ and took $K = L = 9$. The same MSCN process is also applied to each spatial frame F^t, yielding normalized coefficients $\hat{F}^t$. Here after we will drop the temporal index t and the spatial indices $\mathbf{x}$ in (1)-(5)

2) Statistical Characterization: To characterize the statistical features from the normalized MSCN maps, we use the parametric asymmetric generalized Gaussian distribution (AGGD) [43]:

$$f(x; \mu, \gamma, \beta_l, \beta_r) = \begin{cases} \dfrac{\gamma}{(\beta_l + \beta_r)\Gamma(\frac{1}{\gamma})} \exp((-(\dfrac{-x}{\beta_l})^\gamma)), & \forall\, x \le 0 \\ \dfrac{\gamma}{(\beta_l + \beta_r)\Gamma(\frac{1}{\gamma})} \exp((-(\dfrac{-x}{\beta_r})^\gamma)), & \forall\, x \ge 0. \end{cases} \tag{6}$$

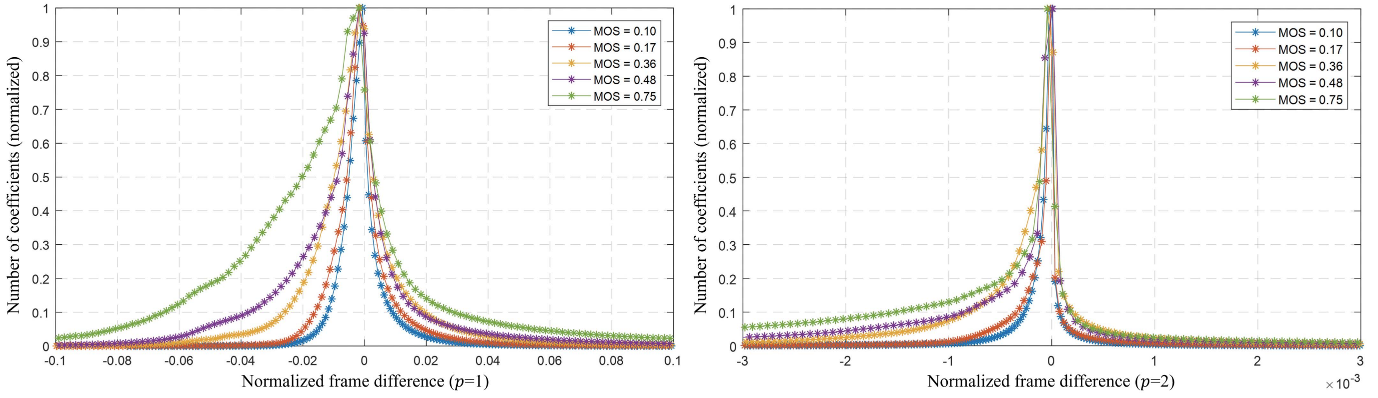
To estimate the parameters of the AGGD ($\mu, \gamma, \beta_l, \beta_r$), we fit it to the histograms of MSCN coefficients over T frames on each VR video sample (and on frames, displaced frame differences, and rotated frame differences), using the popular moment-matching based approach in [43].

In this way, we extract various statistical features on five frame difference maps: ΔF_p, ΔR_p, ΔZ_p, ΔH_p, and ΔV_p (all of which are functions of ($\mathbf{x}$,t), with notation dropped for



(a) (b) (c) (d) (e)

Fig. 8. Depiction of modes and reconstructed maps; (a) nine direction modes according to cardinal angle, (b) reconstructed roll motion maps, (c) zoom in/out motion maps, (d) linear horizontal motion maps, (e) linear vertical motion maps.



(a) (b)

Fig. 9. Histograms of processed frame difference maps ((a) $p = 1$ and (b) $p = 2$) computed on the "Boat" VR contents for various sickness MOS levels.

brevity), where the norm index in 1 is either $p = 1$ or $p = 2$. Specifically, after applying MSCN normalization, 11 feature maps are obtained for each frame on the spatio-temporal frame difference map $\Delta\hat{F}_p$ and the normalized rotational frame difference maps ($\Delta\hat{R}_p$, $\Delta\hat{Z}_p$, $\Delta\hat{H}_p$ and $\Delta\hat{V}_p$) using two norms ($p \in \{1, 2\}$) in addition to the normalized spatial frame $\hat{F}$. Finally, we extract statistical features (μ, γ, β_l, β_r) from each spatial feature map, denoting them collectively as ξ^k, $k \in \{1, \ldots, 4\}$, and likewise extract spatio-temporal features ϕ_p^l, $l \in \{1, \ldots, 4\}$ on normalized frame differences, and rotational features δ_p^m, $m \in \{1, \ldots, 16\}$ on normalized, rotated frame differences.

Fig. 9 depicts the histogram of MSCN coefficients of ΔF_p for p-norm parameters (a) $p = 1$ and (b) $p = 2$ on the 'Boat' VR sequence, along with and the corresponding sickness mean opinion scores (MOS). As shown in the figure, the two different norm parameters result in different histogram shapes. As the MOS increases, the histograms widen.

Furthermore, Fig. 10 shows 3D scatter plots of the (logarithms of the) extracted parameters γ, β_l, β_r obtained by the fitting process in (6), on the five different frame difference maps (ΔF, ΔR, ΔZ, ΔH and ΔV), on all the VR sequence samples in the VR-SP database. Each sample is colored as closer to purple as the feelings of VR sickness MOS increased, and towards yellow with decreased feelings of VR sickness. It may be seen that the samples vary with the degree of sickness, suggesting that the extracted features can play an important role in predicting VR sickness.

E. Exceptional Motion Feature

1) Instantaneous vs. Averaged Temporal Correlation: Prior VR sickness studies [44], [45], have suggested that the visual-vestibular conflict is strongly affected by exceptional motions such as rapid turning movements, *e.g.,* a user experiencing VR content of driving a racing car on a road with many roadblocks. An intuitive way to model exceptional motion is to compare temporal motion flow to spatially local motion flow. To do this, we utilize instantaneous vs. local temporal correlation feature, whereby we measure the correlations between features drawn from instantaneous normalized coefficients (*i.e.,* spatio-temporal features ϕ_p^l) and locally averaged temporal normalized coefficients.

2) Temporal Normalization: We slightly modify the statistics of the MSCN normalization (3) to characterize averaged temporal statistics over multiple frames. Using the subscript

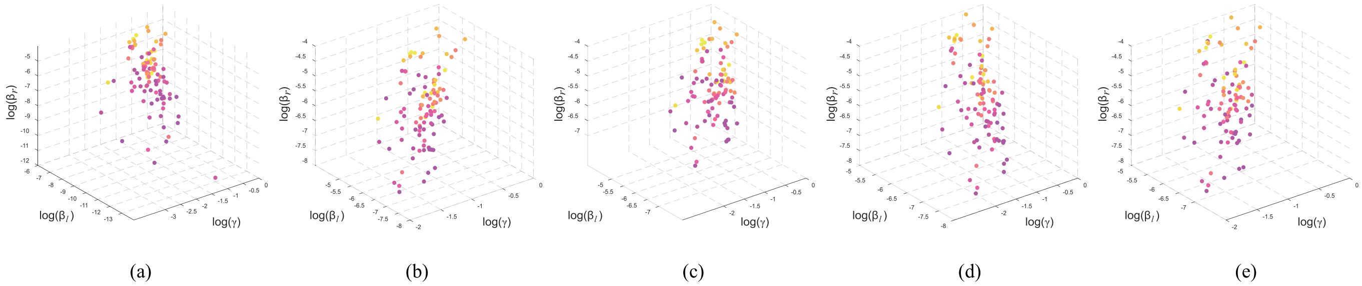


Fig. 10. 3D scatter plots of (logarithms of) shape, left scale and right scale obtained by fitting the AGGD model to (a) the spatio-temporal frame differences, (b) the roll motion frame differences, (c) the zoom-in/out motion frame differences, (d) the horizontal motion frame differences and (e) the vertical motion frame differences, using all the VR video samples from the VR-SP database. Purple colors indicate MOS associated with greater feelings of sickness, while yellow samples indicate MOS values associated with lesser feelings of sickness.

'MF' to denote multi-frame, define

$$\Delta\mu^t_{\mathrm{MF}}(\mathbf{x}) = \sum_{k=-K}^{K}\sum_{l=-L}^{L}\sum_{m=-M}^{M} w_{k,l,m}\,\Delta F^t(\mathbf{x}_1+k,\mathbf{x}_2+l,t+m) \tag{7}$$

and

$$\Delta\sigma^t_{\mathrm{MF}}(\mathbf{x}) = \left(\sum_{k=-K}^{K}\sum_{l=-L}^{L}\sum_{m=-M}^{M} w_{k,l,m}[\Delta F^t(\mathbf{x}_1+k,\mathbf{x}_2+l,t+m) - \Delta\mu^t_{\mathrm{inter}}(\mathbf{x})]^2\right)^{\frac{1}{2}} \tag{8}$$

where $\Delta F^t(\mathbf{x})$ is as before, over a set of consecutive frame time samples $t \in [1, T]$. Then the multi-frame MSCN normalization $\Delta\hat{F}^t_{\mathrm{MF}}$ is computed by using $\Delta\mu_{\mathrm{MF}}$ and $\Delta\sigma_{\mathrm{MF}}$ as in (3). We compute all the same parameters as before, and denote the overall feature vector as $\boldsymbol{\Psi}_p = \{\psi^l_p; l = 1, \ldots, 4\}$.

3) Correlation Between Instantaneous and Averaged Coefficients: The spatio-temporal features ψ^l_p, $l \in \{1, \ldots, 4\}$ of $\Delta\hat{F}_{\mathrm{MF}}$ are obtained by the same AGGD fitting procedure as in Section IV-D. The correlation feature $ß_p$ between the instantaneous and averaged feature vectors ($\boldsymbol{\Phi}_p = \{\phi^l_p\}$ and $\boldsymbol{\Psi}_p = \{\psi^n_p\}$) is

$$ß_p = \sum_l \left(\frac{\phi^l_p - \mu(\boldsymbol{\Phi}_p)}{\sigma(\boldsymbol{\Phi}_p)}\right)\left(\frac{\boldsymbol{\psi}^l - \mu(\boldsymbol{\Psi}_p)}{\sigma(\boldsymbol{\Psi}_p)}\right), \tag{9}$$

where p is the norm parameter $p \in \{1, 2\}$, and $\mu(\cdot)$ and $\sigma(\cdot)$ denote the average and variance of each feature vector, respectively.

Fig. 11 plots the histograms of the correlation feature $ß_1$ for low sickness MOS (MOS < 0.2) and high sickness MOS (MOS > 0.7) instances on the VR-SP database. As may be seen, the correlations are widely distributed for low sickness VR sequences, as compared to high sickness VR sequences.

V. APPLICATION: VR PRESENCE ASSESSOR

As mentioned earlier in Section III (also shown in Fig. 5), there are strong non-linear trade-offs between VR sickness and VR presence. Therefore, we use the predicted VR sickness scores produced by VRSA as a feature, to help estimate the degree of VR presence. As shown in Fig. 5, the VR presence scores (vertical axis) are distributed with a large variance. Therefore, to accurately estimate VR presence, a number of presence-directed features are used to drive the VRPA model [46], [47]: a visual activity feature, and several content features (luminance gradient, color gradient, luminance saturation, and color saturation).

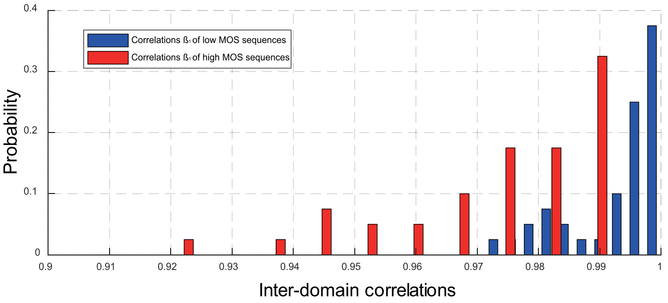


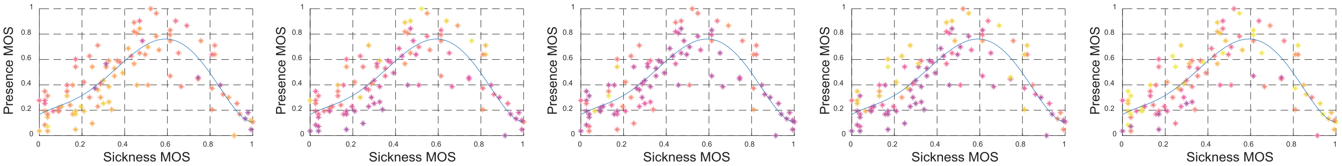
Fig. 11. Probability distributions of correlation feature $ß_1$ on VR sequences having low sickness MOS (MOS < 0.2) and having high sickness MOS (MOS > 0.7) on the VR-SP database.

A. Additional Features for VR Presence

Fig. 12 shows the same scatter diagram as Fig. 5, but with each plotted point color-coded by the values of each of the additional features. Specifically, in Fig. 12 (a)-(e), the color of each sample indicates larger feature values close to yellow, and lower feature values are closer to purple. It may be seen that the distribution of each feature is nicely divided in the vertical direction. Each feature is calculated as follows:

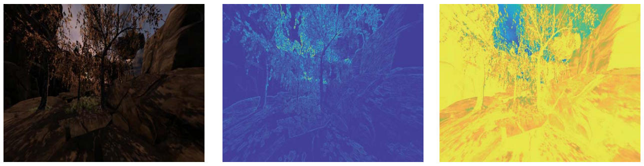
1) Visual Activity: Visual activity has previously been used to quantify visual comfort, preference and presence based on statistical analyses of visual content feature maps [46], [47]. To calculate visual activity, first compute the content feature maps (luminance/color gradient and saturation). Then, the normalized luminance/color gradient map is obtained as

$$\mathcal{G}_l = \frac{1}{\mathcal{G}^m_l}\sqrt{\frac{\partial I^2}{\partial u} + \frac{\partial I^2}{\partial u}} \tag{10}$$



(a) Visual activity (b) Luminance gradient (c) Color gradient (d) Luminance saturation (e) Color saturation

Fig. 12. Distributions of VR presence feature values coded by color in scatter plot of Fig. 5; for features: (a) visual activity, (b) luminance gradient, (c) color gradient, (d) luminance saturation, and (e) color saturation. Yellow colors indicate larger values, while purple colors indicate lower values.



(a) (b) (c)

Fig. 13. Examples of content feature maps. (a) Original "Horse" scene from the VR-SP database, and its (b) normalized luminance gradient map and (c) normalized saturation maps.

and

$$\mathcal{G}_c = \frac{1}{2\mathcal{G}^m_{c,a}}\sqrt{\frac{\partial C_a^2}{\partial u} + \frac{\partial C_a^2}{\partial u}} + \frac{1}{2\mathcal{G}^m_{c,b}}\sqrt{\frac{\partial C_b^2}{\partial u} + \frac{\partial C_b^2}{\partial u}} \tag{11}$$

where $\mathcal{G}_l$ and $\mathcal{G}_c$ are the normalized luminance and color gradients, and u and v are horizontal and vertical indices, respectively. I is the luminance map, and $\mathcal{G}^m_l$ is the maximum luminance gradient of I, which is used as a normalization factor. C_a and C_b are color maps computed after converting the images into the perceptually uniform CIELab color space [48]. $\mathcal{G}^m_{c,a}$ and $\mathcal{G}^m_{c,b}$ are the maximum color gradients of C_a and C_b, which are used for normalization. Fig. 13 (b) shows an exemplar luminance gradient map. In addition, the normalized luminance/color saturation is given by

$$\mathcal{S}_l = \frac{127 - I(u,v)}{127} \tag{12}$$

and

$$\begin{cases} \mathcal{S}_{ca} = \dfrac{127 - C_a(u,v)}{127}, \\ \mathcal{S}_{cb} = \dfrac{127 - C_b(u,v)}{127} \end{cases} \tag{13}$$

where $\mathcal{S}_{ca}$ and $\mathcal{S}_{cb}$ are the normalized color saturations on color spaces C_a and C_b, respectively. Then the normalized color saturation is represented by $\mathcal{S}_c = \mathcal{S}_{ca} \times \mathcal{S}_{cb}$. When the luminance/color saturation values are towards brighter or darker, then the saturation values approach 1. Fig. 13 (c) shows an example of a normalized luminance saturation map.

A discrete wavelet transform is performed on each normalized luminance/color gradient and saturation map. Then, the histogram of the wavelet coefficients of each subband are fitted with a generalized Gaussian distribution (GGD) [49]. Let $\gamma_{k,i}$ denote the GGD shape parameter of the i^{th} wavelet sub-band of the k^{th} feature map (k $\in \{\mathcal{G}_l, \mathcal{G}_c, \mathcal{S}_l$ and $\mathcal{S}_c\}$).

We use the shape parameter as a major descriptor of visual activity, since it captures the distribution of energy across the wavelet-transformed feature maps [47]. Using the shape parameter, the visual activity $\mathcal{A}_{k,i}$ of the i^{th} sub-band of the k^{th} feature map is defined as:

$$\mathcal{A}_{k,i} = \begin{cases} 1 - \dfrac{1}{1+\exp\{-2(\frac{\gamma_{k,i}-c_k}{c_{k,D}})\}}, & \gamma_{k,i} < c_k \\ 1 - \dfrac{1}{1+\exp\{-2(\frac{\gamma_{k,i}-c_k}{c_{k,U}})\}}, & \text{otherwise} \end{cases} \tag{14}$$

where the decomposition level is set to 2 while the number of sub-bands i is set to 7; c_k, $c_{k,D}$ and $c_{k,U}$ are normalization factors on the k^{th} feature map. Since the range of visual activity varies with the feature maps, the normalization factors c_k, $c_{k,D}$ and $c_{k,U}$ are applied on the k^{th} feature map, where c_k is the reference operating point of the k^{th} feature map, and $c_{k,D}$ and $c_{k,U}$ are the trailing and leading edges of the region over which (14) is approximately linear [47]. In our experiments, the normalization factors were obtained by in [50],

$$\begin{cases} (c_{\mathcal{G}_l}, c_{\mathcal{G}_{l,D}}, c_{\mathcal{G}_{l,U}}) = (0.964, 0.853, 1.554), \\ (c_{\mathcal{G}_c}, c_{\mathcal{G}_{c,D}}, c_{\mathcal{G}_{c,U}}) = (0.964, 0.853, 1.554), \\ (c_{\mathcal{S}_l}, c_{\mathcal{S}_{l,D}}, c_{\mathcal{S}_{l,U}}) = (1.312, 0.677, 1.388), \\ (c_{\mathcal{S}_c}, c_{\mathcal{S}_{c,D}}, c_{\mathcal{S}_{c,U}}) = (1.172, 0.398, 1.075). \end{cases}$$

Finally, we calculate the visual activity of each feature map by aggregating them over all sub-bands:

$$\mathcal{A} = \sum_k w_k \mathcal{A}_k, \tag{15}$$

where $\mathcal{A}_k$ is the activity of the k^{th} feature map having N_{sub} sub-bands: $\mathcal{A}_k = \sum_i \frac{\mathcal{A}_{k,i}}{N_{sub}}$. Here, w_k is an empirical weight on each feature map.

2) Content Features: We summarize the content feature maps (*CFs*) using three different pooling methods. The first is the mean value of (each) feature map, while the other two features are computed by calculating the upper and lower s-percentiles % (mean of upper/lower s% of each feature map,), respectively. This yields 12 pooled CF features.

B. Overall Feature Configurations

We use the features defined above in Sections IV-V and the VR sickness/presence MOS scores to train a support vector regression (SVR). The extracted VRSA features can be categorized as spatial features (*SF*), spatio-temporal features (*STF*), rotational features (*RF*) and exceptional motion features (*EMF*). For VRPA, there are sickness features (SCF), visual activity features (AF), and content features (CF). Overall, there are 46 VR sickness features and 14 VR presence features.

VI. EXPERIMENTAL RESULTS

A. Experimental Setup

To validate the performance of VRSA and VRPA, we employed two standard measures: Pearson's linear correlation coefficient (PLCC) [52], and Spearman's rank-order correlation coefficient (SROCC). A value close to 1 for SROCC and PLCC indicates better performance. We followed the validation strategy in [53]–[57]: first, we randomly divided the VR video database (reference and training) into two content-separated subsets (80% for training and 20% for testing). An SVR was utilized as the regression tool, since it has demonstrated excellent performance on other high-dimensional regression problems, such as 3D discomfort assessment and visual preference prediction [46]. To implement the SVR, we used the libSVM package [58] with the radial basis kernel.

The correlation results that we report are the median correlation over 2000 iterations of randomly dividing the training and testing sets to eliminate any biases (cross-validation).

B. Dataset

We benchmarked the VRSA prediction model on two different datasets: ETRI-VR [30] and the new VR-SP dataset. The VRPA prediction model was tested on the VR-SP dataset. The ETRI-VR database includes 2 reference VR contents with 52 slightly modified scenarios, along with corresponding subjective sickness scores [30]. As mentioned earlier, the VR-SP dataset includes 10 reference VR contents and 100 variations of the reference VR contents with their corresponding sickness and presence MOS. All sickness MOS values in these databases are scaled to [0, 1], where 1 represents to the extreme discomfort.

C. VRSA Prediction Results

We compared the predictions of VRSA against five VR sickness assessment models: the optical flow-based method in [51], VRSP [7], Kim *et al.* [30], Kim *et al.* [8] and VRSA-NET [9]. For the optical flow-based method, the model was implemented using an estimated average optical flow magnitude [51]. For Kim *et al.* [8] and VRSA-NET [9], the experimental setup was the same as in the original work. In the case of Kim *et al.* [30], since our database does not include brain signals, we only utilized visual features from the combined CNN-RNN network.

1) Benchmark Results: Table III tabulates the performance comparison over the tested models on the VR-SP and ETRI-VR databases. When measured on all databases, the standard deviations of PLCC and SROCC after 2000 trials were less than ~0.03. As shown in Table, VRSA delivered significantly better predictive performance than the other models in terms of both correlation and reliability.

TABLE III

PLCC AND SROCC COMPARISON ON TWO VR SICKNESS ASSESSMENT DATABASES

Method	VR-SP (*proposed*)		ETRI-VR	
	PLCC	SROCC	PLCC	SROCC
Optical flow-based [51]	0.565	0.575	0.702	0.690
VRSP [7]	0.683	0.756	0.725	0.710
Kim *et al.* [30]	0.836	0.833	0.689	0.690
Kim *et al.* [8]	0.797	0.712	0.661	0.646
VRSA-NET [9]	0.843	0.814	0.745	0.712
VRSA (*proposed*)	**0.903**	**0.894**	**0.789**	**0.721**

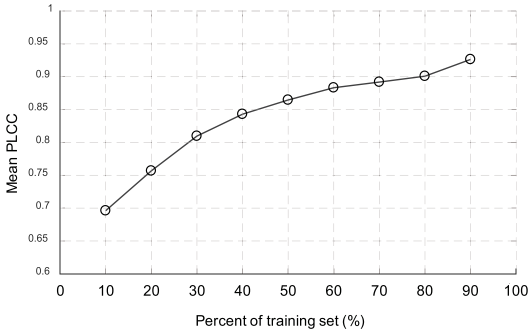


Fig. 14. Mean PLCC performance of VRSA against the percentage of the overall dataset that was used for training.

We also conducted a t-test of statistical significance on the SROCC values over 50 trials of all pairs of benchmark models. Table VI shows the results of the t-tests on the VR-SP database. The symbols "1", "0" and "−1" indicate that the performance of the model in the row is statistically better, indistinguishable, or worse, respectively, than the compared methods in the column. We set the confidence level to be 95% (*i.e.,* significance is determined when the p-value is less than 0.05). As shown in the results, VRSA was more predictive of VR sickness than the other benchmark models with statistical significance.

2) Performance on Individual Motion Types: Since the VR-SP database contains a variety of motion types, we also tested the model according to these types. Table V reports the SROCC of the compared VRSA algorithms on the VR-SP database against motion type. As shown in the table, VRSA delivered the best performance on both motion types.

3) Dependency on Train/Test Proportion: In order to study the degree on the performance of the model, we measured the mean values of the PLCCs over 2000 trials as a function of the training set percentage as it ranged from 10% to 90% in 10% increments. Fig. 14 shows as the percentage of the training set increased from 50% to 90%, the performance difference varied less than 10%.

4) Feature Ablation Study: We also studied the performances of the feature groups and combinations of feature groups. Table IV shows the LCC and SROCC for each feature group of combinationsacross 2000 train-test trials. Specifically, we tested *SF*, *STF*, *RF*, *EMF*, and combinations (*SF*+*STF*), (*SF*+*RF*), (*SF*+*EMF*), (*SF*+*STF*+*RF*) and (*SF*+*RF*+*EMF*).

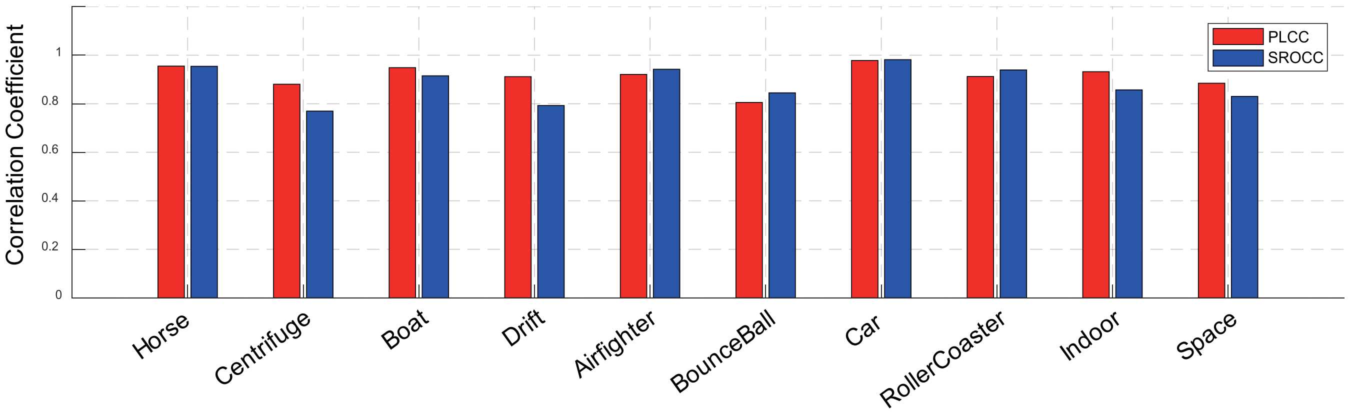


Fig. 15. Plots of PLCC and SROCC on each reference VR video: 'Horse', 'Centrifuge', 'Boat', 'Drift', 'AirFighter', 'BounceBall', 'Car', 'RollerCoaster', 'Indoor','Space'.

TABLE IV

PLCC AND SROCC OF FEATURES AND THEIR COMBINATIONS OVER 2000 TRAIN-TEST TRIALS ON THE VR-SP DATABASE

Methods	PLCC	SROCC	Methods	PLCC	SROCC
SF	0.328	0.286	*SF+STF*	0.843	0.825
STF	0.770	0.754	*SF+RF*	0.880	0.873
RF	0.820	0.792	*SF+EMF*	0.721	0.699
EMF	0.719	0.677	*SF+STF+RF*	0.892	0.876
SF+RF+EMF	0.888	0.885	**VRSA** (*full*)	**0.903**	**0.894**

TABLE V

SROCC OF THE VRSA MODELS AGAINST MOTION TYPES ON THE VR-SP DATABASE

Methods	Optical flow-based [51]	VRSP [7]	Kim *et al.* [30]	Kim *et al.* [8]	VRSA-NET [9]	VRSA
Motion Type 1	0.537	0.753	0.842	0.814	0.821	**0.923**
Motion Type 2	0.675	0.876	0.874	0.876	0.883	**0.964**

TABLE VI

RESULTS OF ONE SIDED T-TEST PERFORMED ON THE SRCC RESIDUALS ON THE VR-SP DATABASE. THE SIGNIFICANCE LEVEL WAS SET TO 95%

	optical flow-based [51]	VRSP [7]	Kim *et al.* [30]	Kim *et al.* [8]	VRSA-NET [9]	VRSA
Optical flow-based [51]	0	-1	1	-1	1	1
VRSP [7]	1	0	1	-1	1	1
Kim *et al.* [30]	-1	1	0	1	-1	1
Kim *et al.* [8]	1	1	-1	0	1	1
VRSA-NET [9]	1	-1	1	-1	0	1
VRSA	-1	-1	-1	-1	-1	0

TABLE VII

PERFORMANCE OF VRSA AS A FUNCTION OF FILTER WINDOW SIZE ON THE VR-SP DATABASE

Gaussian Filter Size	PLCC	SROCC
7	0.812	0.834
11	0.858	0.853
15	**0.903**	**0.894**
19	0.873	0.820

Since the spatial feature *SF* is not a major contributor to VR sickness, its correlation against MOS is only about 0.3. However, the frame difference map-based statistical features (*STF*, *RF* and *EMF*) delivered reliable predictive performance. In particular, the correlation obtained by *RF* were higher than those of the other single features. When *SF* was combined with *STF*, *RF* and *EMF*, the correlation *SF+RF+EMF* delivered even higher SROCC performance, indicating that EMF efficiently contributes to performance as well. Overall, when all the features were used together, VRSA much better performance than the other models.

D. Content Dependence

We also explored the way that VRSA behaves across diverse video contents. Fig. 15 shows the performance of VRSA over the 2000 trials on each of the reference VR videos in the VR-SP database. Although the correlation it attained was slightly better on some scenarios than the others, the variation in performance lay within a fairly small range.

E. Effect of Model Parameters

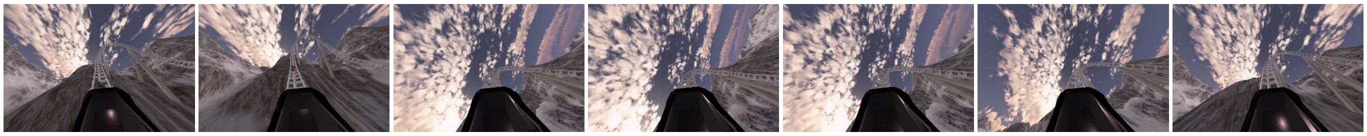
To study model behavior against the parameters of VRSA, we tabulated performance while varying the filter window sizes. Table VII shows performances for four filter windows on the VR-SP database. It can be seen that the algorithm performed best for a Gaussian filter window of size 15.

F. VRPA Prediction Results

To evaluate the performance of VRPA, we used the subjective presence assessment in Section III and the additional features defined in Section IV. Similar to the analysis of VRSA, SROCC and PLCC were used to measure performance. The overall experimental protocol was as described in Section VI-A.

(a) Sickness: 0.972 (1.000) Presence: 0.233 (0.106)

(b) Sickness: 0.773 (0.896) Presence: 0.375 (0.298)



(c) Sickness: 0.552 (0.521) Presence: 0.728 (1.000)

(d) Sickness: 0.143 (0.083) Presence: 0.291 (0.191)

Fig. 16. Example of VR sequences and their predicted (MOS scaled to [0, 1]) sickness/presence scores on the VR-SP database. (a) ‘Horse-7’ sequence, (b) ‘Centrifuge-6’ sequence, (c) ‘RollerCoaster-2’ sequence, and (d) ‘Space-1’ sequence.

TABLE VIII
PLCC AND SROCC OF VRPA ON THE VR-SP DATABASE

Methods	PLCC	SROCC
SCF	0.694	0.678
VAF	0.432	0.461
CF	0.412	0.346
VRPA (*full*)	**0.747**	**0.753**

1) Feature Ablation Study: We studied the individual performances of the proposed VR presence feature sets and combinations of them. Table VIII shows the LCC and SROCC obtained for each feature set and their combinations across 2000 train-test trials. We tested four different models: *SCF*, *VAF*, *CF* and *VRPA (full)*.

As discussed in Section V, the VR sickness feature is a good prediction of the sense of VR presence, and it delivers a reliable correlation of ∼0.69. When all of the features are used together, VRPA attains a top correlation performance about ∼0.75. These results infer the necessity of considering all of the sickness, visual activity and content features simultaneously to evaluate VR presence accurately.

2) Predicted Results: To determine whether VR sickness is related to presence, four sampled VR sequences and their predicted sickness/presence scores (MOS scaled to [0, 1]) are illustrated in Fig. 16. Overall, it can be seen that the predicted scores are close to MOS. Also, the sequences with high VR sickness scores, such as shown in Figs. 16 (a) and (b), have relatively low VR presence scores. Similarly, as shown in Fig. 16 (d), when there is a very low level of VR sickness, VR presence becomes quite reduced. However, in Fig. 16 (c) exemplifying an appropriate level of VR sickness, the sense of presence is relatively maximized. Broadly, it may be seen that there is a strong relationship between VR sickness and VR presence, and our prediction models agree with that observation.

VII. CONCLUSION

Recently, virtual experience technology has improved remarkably, raising the importance of VR sickness/presence predictors in commercial HMD products. Reliable QoE assessment algorithms could help provide users with more realistic and comfortable visual experiences when using VR. In the VR industry, significant efforts have been made to improve software/hardware techniques to better deliver viewers’ visual satisfaction. Here, we described a new large-scale database of VR sickness/presence and that we created, and used it to formulate new VR prediction models. Our simulation results show that our proposed models can autonomously and effectively predict VR sickness/presence, achieving much better performance than conventional algorithms. We expect that the proposed models could be applied in a variety of applications, such as VR content creation, VR broadcasting, and VR gaming.

REFERENCES

[1] W. Kim *et al.*, “Modern trends on quality of experience assessment and future work,” *APSIPA Trans. Signal Inf. Process.*, vol. 8, no. 23, pp. 1–19, Oct. 2019.

[2] L. Rebenitsch and C. Owen, “Review on cybersickness in applications and visual displays,” *Virtual Reality*, vol. 20, no. 2, pp. 101–125, Jun. 2016.

[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, Jul. 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. 2015.

[5] 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,” *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 46, no. 1, pp. 142–153, Mar. 2004.

[6] T. Brandt *et al.*, “Visual-vestibular and visuovisual cortical interaction,” *Ann. New York Acad. Sci.*, vol. 956, no. 1, pp. 230–241, Apr. 2002.

[7] J. Kim, W. Kim, S. Ahn, J. Kim, and S. Lee, “Virtual reality sickness predictor: Analysis of visual-vestibular conflict and VR contents,” in *Proc. 10th Int. Conf. Qual. Multimedia Exper. (QoMEX)*, May 2018, pp. 1–6.

[8] 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. 23rd ACM Symp. Virtual Reality Softw. Technol.*, Nov. 2017, p. 36.

[9] 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.*, vol. 28, no. 4, pp. 1646–1660, Apr. 2019.

[10] 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, 2005.

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

[12] 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.

[13] J.-R. Chardonnet, M. A. Mirzaei, and F. Mérienne, “Features of the postural sway signal as indicators to estimate and predict visually induced motion sickness in virtual reality,” *Int. J. Hum.–Comput. Interact.*, vol. 33, no. 10, pp. 771–785, Oct. 2017.

[14] T. Brandt, J. Dichgans, and E. Koenig, “Differential effects of central versus peripheral vision on egocentric and exocentric motion perception,” *Exp. Brain Res.*, vol. 16, no. 5, pp. 476–491, Mar. 1973.

[15] R. Hosman, F. Cardullo, and J. Bos, “Visual-vestibular interaction in motion perception,” in *Proc. AIAA Model. Simul. Technol. Conf.*, Aug. 2011, p. 6425.

[16] H. R. Sheikh and A. C. Bovik, “Image information and visual quality,” *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[17] A. K. Moorthy and A. C. Bovik, “Blind image quality assessment: From natural scene statistics to perceptual quality,” *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.

[18] M. A. Saad, A. C. Bovik, and C. Charrier, “Blind image quality assessment: A natural scene statistics approach in the DCT domain,” *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.

[19] A. Mittal, A. K. Moorthy, and A. C. Bovik, “No-reference image quality assessment in the spatial domain,” *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[20] A. Mittal, M. A. Saad, and A. C. Bovik, “A completely blind video integrity oracle,” *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 289–300, 2015.

[21] W. A. Ijsselsteijn, H. de Ridder, J. Freeman, and S. E. Avons, “Presence: Concept, determinants, and measurement,” *Proc. SPIE*, vol. 3959, pp. 520–529, Jun. 2000.

[22] W. IJsselsteijn, H. D. Ridder, J. Freeman, S. E. Avons, and D. Bouwhuis, “Effects of stereoscopic presentation, image motion, and screen size on subjective and objective corroborative measures of presence,” *Presence, Teleoperators Virtual Environ.*, vol. 10, no. 3, pp. 298–311, Jun. 2001.

[23] H. Oh and S. Lee, “Visual presence: Viewing geometry visual information of UHD S3D entertainment,” *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3358–3371, Jul. 2016.

[24] S. Bangay and L. Preston, “An investigation into factors influencing immersion in interactive virtual reality environments,” *Stud. Health Technol. Informat.*, pp. 43–51, 1998.

[25] S. Weech, S. Kenny, and M. Barnett-Cowan, “Presence and cybersickness in virtual reality are negatively related: A review,” *Frontiers Psychol.*, vol. 10, p. 158, Feb. 2019.

[26] 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*, Mar. 2002, pp. 164–171.

[27] J. D. Prothero, “The role of rest frames in vection, presence and motion sickness,” Univ. Washington, Seattle, WA, USA, Tech. Rep., 1998.

[28] M. M. Knight and L. L. Arns, “The relationship among age and other factors on incidence of cybersickness in immersive environment users,” in *Proc. ACM SIGGRAPH Res. Posters - SIGGRAPH*, 2006, p. 196.

[29] I. Milleville-Pennel and C. Charron, “Do mental workload and presence experienced when driving a real car predispose drivers to simulator sickness? An exploratory study,” *Accident Anal. Prevention*, vol. 74, pp. 192–202, Jan. 2015.

[30] 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 *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 10580–10589.

[31] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document R. I.-R. BT.500, 2002.

[32] *Subjective Methods for the Assessment of Stereoscopic 3Dtv Systems*, document R. I.-R. BT.2021, 2012.

[33] H. Ujike, T. Yokoi, and S. Saida, “Effects of virtual body motion on visually-induced motion sickness,” in *Proc. 26th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2004, pp. 2399–2402.

[34] B. G. Witmer and M. J. Singer, “Measuring presence in virtual environments: A presence questionnaire,” *Presence, Teleoperators Virtual Environ.*, vol. 7, no. 3, pp. 225–240, Jun. 1998.

[35] N. McHugh, S. Jung, S. Hoermann, and R. W. Lindeman, “Investigating a physical dial as a measurement tool for cybersickness in virtual reality,” in *Proc. 25th ACM Symp. Virtual Reality Softw. Technol.*, Nov. 2019, pp. 1–5.

[36] X. Li, Q. Guo, and X. Lu, “Spatiotemporal statistics for video quality assessment,” *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3329–3342, Jul. 2016.

[37] H. Ojanen, “Automatic correction of lens distortion by using digital image processing,” Rutgers Univ., Dept. Math., Camden, NJ, USA, Tech. Rep., 1999.

[38] R. Soundararajan and A. C. Bovik, “Video quality assessment by reduced reference spatio-temporal entropic differencing,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 4, pp. 684–694, Apr. 2013.

[39] M. E. McCauley, J. W. Royal, C. D. Wylie, J. F. O’Hanlon, and R. R. Mackie, “Motion sickness incidence: Exploratory studies of habituation, pitch and roll, and the refinement of a mathematical model,” Canyon Research Group Inc Goleta Ca Human Factors Research Div, Tech. Rep., 1976.

[40] H. V. Howarth and M. J. Griffin, “Effect of roll oscillation frequency on motion sickness,” *Aviation, Space, Environ. Med.*, vol. 74, no. 4, pp. 326–331, 2003.

[41] Z. Sinno and A. C. Bovik, “Spatio-temporal measures of naturalness,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 1750–1754.

[42] D. L. Ruderman, “The statistics of natural images,” *Netw., Comput. Neural Syst.*, vol. 5, no. 4, pp. 517–548, 1994.

[43] N.-E. Lasmar, Y. Stitou, and Y. Berthoumieu, “Multiscale skewed heavy tailed model for texture analysis,” in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 2281–2284.

[44] 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, p. 472, Apr. 2015.

[45] A. Mazloumi Gavgani, D. M. Hodgson, and E. Nalivaiko, “Effects of visual flow direction on signs and symptoms of cybersickness,” *PLoS ONE*, vol. 12, no. 8, Aug. 2017, Art. no. e0182790.

[46] H. Kim, S. Ahn, W. Kim, and S. Lee, “Visual preference assessment on ultra-high-definition images,” *IEEE Trans. Broadcast.*, vol. 62, no. 4, pp. 757–769, Dec. 2016.

[47] K. Lee, A. K. Moorthy, S. Lee, and A. C. Bovik, “3D visual activity assessment based on natural scene statistics,” *IEEE Trans. Image Process.*, vol. 23, no. 1, pp. 450–465, Jan. 2014.

[48] K. T. Mullen, “The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings.,” *J. Physiol.*, vol. 359, no. 1, pp. 381–400, Feb. 1985.

[49] K. Sharifi and A. Leon-Garcia, “Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 1, pp. 52–56, 1995.

[50] R. Cappelli, D. Maio, and D. Maltoni, “Combining fingerprint classifiers,” in *Proc. Int. Work. Multi. Class. Syst. (MCS)*. Berlin, Germany: Springer, 2000, pp. 351–361.

[51] B. K. P. Horn and B. G. Schunck, “Determining optical flow,” *Artif. Intell.*, vol. 17, nos. 1–3, pp. 185–203, Aug. 1981.

[52] *Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II*, VQEG, Boulder, CO, USA, 2003.

[53] J. Kim and S. Lee, “Fully deep blind image quality predictor,” *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 1, pp. 206–220, Feb. 2017.

[54] J. Kim, H. Zeng, D. Ghadiyaram, S. Lee, L. Zhang, and A. C. Bovik, “Deep convolutional neural models for picture-quality prediction: Challenges and solutions to data-driven image quality assessment,” *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 130–141, Nov. 2017.

[55] 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 *Proc. Eur. Conf. Comput. Vis.(ECCV)*, vol. 2018, pp. 219–234.

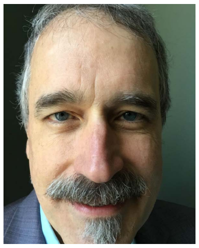
[56] W. Kim, A.-D. Nguyen, S. Lee, and A. C. Bovik, “Dynamic receptive field generation for full-reference image quality assessment,” *IEEE Trans. Image Process.*, vol. 29, pp. 4219–4231, 2020.

[57] J. Kim, A.-D. Nguyen, and S. Lee, “Deep CNN-based blind image quality predictor,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 1, pp. 11–24, Jan. 2018.

[58] C.-C. Chang and C.-J. Lin, “LIBSVM: A library for support vector machines,” *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, 2011.

Woojae Kim received the B.S. degree in electronic engineering from Soongsil University, Seoul, South Korea, in 2015. He is currently pursuing the M.S. and Ph.D. degrees with Multidimensional Insight Laboratory, Yonsei University. He was a Research Assistant, under the guidance of Prof. Weisi Lin, with the Laboratory for School of Computer Science and Engineering, Nanyang Technological University (NTU), Singapore, in 2018. His research interests include image and video processing based on the human visual systems, image/video quality assessment, computer vision, and machine learning.

Sanghoon Lee (Senior Member, IEEE) received the B.S. degree from Yonsei University, Seoul, South Korea, in 1989, the M.S. degree from the Korea Advanced Institute of Science and Technology, South Korea, in 1991, and the Ph.D. degree from The University of Texas at Austin, TX, USA, in 2000. From 1991 to 1996, he was with Korea Telecom, South Korea. From 1999 to 2002, he was with Lucent Technologies, NJ, USA. In 2003, he joined the EE Department, Yonsei University, as a Faculty Member, where he is currently a Full Professor. His current research interests include image/video processing, computer vision, and graphics. He received the 2015 Yonsei Academic Award from Yonsei University, the 2012 Special Service Award from the IEEE Broadcast Technology Society, and the 2013 Special Service Award from the IEEE Signal Processing Society. He was the General Chair of the 2013 IEEE IVMSP Workshop, and has served as a steering committee member for IEEE and APISPA conferences. He has been serving as the Chair of the IEEE P3333.1 Quality Assessment Working Group since 2011. He was the IVM Technical Committee Chair of APSIPA from 2018 to 2019, and is a Board of Governors Member of APSIPA in 2020. He was the IEEE IVMSP Technical Committee Member from 2014 to 2019, and has been the IEEE MMSP Technical Committee since 2016. He also served as an Editor for the *Journal of Communications and Networks* from 2009 to 2015, and a Special Issue Guest Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING in 2013. He was an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING from 2010 to 2014. He also served as an Associate Editor from 2014 to 2018, and currently a Senior Area Editor of the IEEE SIGNAL PROCESSING LETTERS.



Alan Conrad Bovik (Fellow, IEEE) is currently the Cockrell Family Regents Endowed Chair Professor at The University of Texas at Austin. His research interests include image processing, digital television, digital streaming video, and visual perception. For his work in these areas, he was a recipient of the 2015 Primetime Emmy Award for Outstanding Achievement in Engineering Development from the Television Academy, the 2017 Edwin H. Land Medal from the Optical Society of America, the 2019 Progress Medal from The Royal Photographic Society, the 2019 IEEE Fourier Award, and the Norbert Wiener Society Award and the Karl Friedrich Gauss Education Award from the IEEE Signal Processing Society. A perennial Web of Science Group Highly-Cited Researcher, he has also received over ten best journal paper awards, including the 2016 IEEE Signal Processing Society Sustained Impact Award. His recent books include *The Essential Guides to Image Processing* and *The Essential Guides to Video Processing*. He co-founded and was the longest-serving Editor-in-Chief of the IEEE TRANSACTIONS ON IMAGE PROCESSING, and also created/chaired the IEEE International Conference on Image Processing, which was first held in Austin, TX, USA, in 1994.