

SACA Net: Cybersickness Assessment of Individual Viewers for VR Content via Graph-based Symptom Relation Embedding

Sangmin Lee¹, Jung Uk Kim¹, Hak Gu Kim², Seongyeop Kim¹,
and Yong Man Ro¹

¹ Image and Video Systems Lab, School of Electrical Engineering, KAIST, Korea
{sangmin.lee, jukim0701, seongyeop, ymro}@kaist.ac.kr

² Signal Processing Lab, Institute of Electrical Engineering, EPFL, Switzerland
{hakgu.kim}@epfl.ch

Abstract. Recently, cybersickness assessment for VR content is required to deal with viewing safety issues. Assessing physical symptoms of individual viewers is challenging but important to provide detailed and personalized guides for viewing safety. In this paper, we propose a novel symptom-aware cybersickness assessment network (SACA Net) that quantifies physical symptom levels for assessing cybersickness of individual viewers. The SACA Net is designed to utilize the relational characteristics of symptoms for complementary effects among relevant symptoms. The proposed network consists of three main parts: a stimulus symptom context guider, a physiological symptom guider, and a symptom relation embedder. The stimulus symptom context guider and the physiological symptom guider extract symptom features from VR content and human physiology, respectively. The symptom relation embedder refines the stimulus-response symptom features to effectively predict cybersickness by embedding relational characteristics with graph formulation. For validation, we utilize two public 360-degree video datasets that contain cybersickness scores and physiological signals. Experimental results show that the proposed method is effective in predicting human cybersickness with physical symptoms. Further, latent relations among symptoms are interpretable by analyzing relational weights in the proposed network.

Keywords: Cybersickness assessment, individual viewer, VR content, physical symptom, symptom relation

1 Introduction

Perceiving virtual reality (VR) content such as 360-degree videos can provide immersive experiences to viewers. With the rapid development of content capturing and displaying devices, VR content increasingly attracts attention in various industry and research fields [14–16]. However, the growth of VR environments accompanies by concerns over the safety of viewing VR content. Several studies reported that viewing VR content could trigger cybersickness with physical

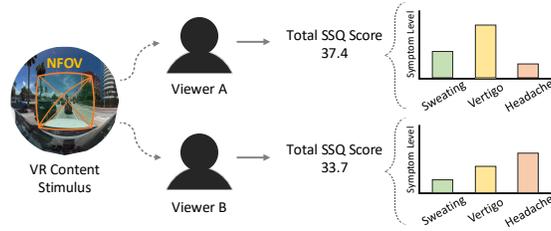


Fig. 1. Viewers watching the same VR content can feel cybersickness differently with distinct symptoms in detail.

symptoms [7,18] : 1) nausea symptoms containing sweating, salivation and, burping, 2) oculomotor symptoms containing visual fatigue and eye strain, and 3) disorientation symptoms containing fullness of the head, dizziness, and vertigo. Such cybersickness is one of the major problems hampering the spread of VR environments. For guiding people to create and view safety content, it is firstly needed to quantify cybersickness level caused by viewing VR content.

When viewers watch VR content, they can feel different cybersickness even for the same content stimulus. As shown in Fig. 1, each viewer can feel cybersickness differently with distinct symptoms in detail. To guide view-safe VR content for specific viewers, it is necessary to quantify detailed cybersickness of individual viewers. Cybersickness assessment for individual viewers needs detailed physical symptoms to provide personalized guides for VR content viewing.

In recent years, VR content-based cybersickness assessment methods have been introduced [20–24]. The content-based methods exploited spatio-temporal features from VR content to quantify cybersickness. These content-based methods did not consider deviations among individuals. Individuals in the same stimulus environment could experience different cybersickness.

There have been clinical studies examining the tendency of physiological responses according to cybersickness [11, 12, 25, 30, 34, 38, 41, 48]. There have been attempts to validate the relationship between physiological responses and cybersickness caused by VR content [11, 13, 25, 33, 43]. Some previous works extract cybersickness-related features from physiological response for predicting cybersickness from VR content [17, 28, 31, 47]. However, since most of them only exploited physiology without stimulus context that affects physiological response predominantly, they did not fully utilize the context for cybersickness. [28] considered stimulus with physiology in evaluating cybersickness. However, it only focuses on predicting total cybersickness levels. Such cybersickness assessment was limited in that it was performed without analysis of physical symptoms.

In this paper, we propose a novel symptom-aware cybersickness assessment network (SACA Net) that predicts the individual viewer’s cybersickness with physical symptoms. The SACA Net quantifies the degree of physical symptoms, which makes it possible to provide more detailed and interpretable information for cybersickness. There were clinical reports about the existence of relationships

among physical symptoms for cybersickness [18, 44]. Considering the relationships, we devise a relational symptom embedding framework that can exploit the relational information among symptoms to assess cybersickness. Thereby, the relevant symptom features complement each other for effectively predicting symptom levels. The SACA Net consists of three parts: a stimulus symptom context guider, a physiological symptom guider, and a symptom relation embedder.

The stimulus symptom context guider is designed to effectively accumulate visual features for encoding symptom factors caused by stimulus environment. Based on neural mismatch theory [40], a sensory mismatch detector in the stimulus symptom context guider extracts mismatch features between target stimulus content and comfort stimulus content that do not induce high-level cybersickness. By exploiting the mismatch features from the sensory mismatch detector, the stimulus symptom context guider extracts symptom group features that represent nausea, oculomotor, and disorientation group factors in context of stimulus.

The physiological symptom guider extracts symptom features from EEG signal. Since EEG contains the most comprehensive information about the nervous system such as vision, movement, and sense [25, 45], we employ EEG as a physiological factor for cybersickness analysis. Considering clinical studies [30, 38, 48] for EEG frequency bands, we design the frequency band attentive encoding for the EEG signal to effectively extract symptom features related to cybersickness.

The symptom relation embedder is designed to refine the symptom features from stimulus and response by embedding relational features. It receives symptom group features and symptom features from the stimulus symptom context guider and the physiological symptom guider, respectively. It learns the latent relations among symptoms in an unsupervised way with graph formulation to effectively predict symptom levels. In addition, we can interpret the relations among symptoms by analyzing relational weights in the proposed network.

We use two public 360-degree video datasets with simulator sickness questionnaire (SSQ) [18] scores and physiological signals. The performances are validated with human cybersickness levels in the assessment datasets.

The major contributions of the paper are as follows.

- We introduce a novel SACA Net that quantifies cybersickness of individuals with physical symptoms by combining content stimulus and physiological response. To the best of our knowledge, it is the first attempt to quantify cybersickness including symptoms of individuals for VR content.
- We propose symptom relation embedding which makes it possible to effectively assess cybersickness of individuals with distinct symptoms. Furthermore, latent symptom relations are interpretable by analyzing relational weights in the proposed network.

2 Related Work

2.1 Cybersickness Assessment for VR Content

VR content-based cybersickness assessment methods have been introduced [20–22, 36]. Kim et al. [20] quantified cybersickness caused by exceptional motions

with a deep generative model. The generative model is trained with normal videos containing non-exceptional motions. Hence, the generative model cannot properly reconstruct videos with exceptional motions that cause cybersickness at testing time. Acquired difference between the original video and the generated video correlated with the degree of cybersickness. In [21], a deep network that consists of a generator and a cybersickness regressor was proposed for quantifying cybersickness. In the model, the difference between the original video and the generated video is regressed to the simulation sickness questionnaires (SSQ) [5] score assessed by subjects. Another study [22] quantified cybersickness considering visual-vestibular conflicts. In the work, to quantify cybersickness, SVR [4] is applied on motion features from visual-vestibular interaction and VR content features. Padmanaban et al. [36] proposed a cybersickness predictor to estimate the nauseogenicity of virtual content. In the model, algorithms based on optical flow methods are employed to compute cybersickness features that primarily focus on disparity and velocity of video content. For assessing cybersickness caused by quality degradation, an objective assessment model considering spatio-temporal inconsistency was proposed [24]. In [23], a deep neural network that exploits cognitive feature regularization was proposed for cybersickness assessment.

However, the aforementioned cybersickness quantification methods do not assess cybersickness of individuals. Individuals in the same environments may experience different cybersickness levels. Compared to these works, the proposed method predicts individual cybersickness by exploiting physiological responses of content viewers to consider the deviation among individuals.

2.2 Physiological Study for Cybersickness

There have been attempts to validate the relationship between cybersickness and physiological responses [11, 13, 25, 31–33, 37, 43, 47]. Kim et al. [25] investigated the characteristic changes of the physiological signals such as EEG, ECG, and GSR while subjects are exposed to VR content. They conducted spectral analysis on each frequency band of EEG signals and validated that specific frequency bands have close correlations with cybersickness. In the case of ECG, they disclosed that the heart period was shorter during the virtual navigation than the baseline period. For GSR, they observed skin conductance level increased during the virtual navigation compared to the baseline. Mawalid et al. [32] attempted to extract EEG statistical feature with PCA for classification in order to investigate cybersickness. Pane et al. [37] adopted on power percentage features extracted from EEG signals to identify cybersickness level. Lin et. al [31] applied linear regression (LR), support vector regression (SVR), and self-organizing neural fuzzy inference network (SOFIN) models on cybersickness-related features extracted from PCA to predict sickness level. Recently, there is a study to predict VR sickness levels by using frequency band power of EEG signal with deep learning structure [17]. There also exists deep learning-based approach utilizing both content analysis and physiology analysis to predict cybersickness [28].

However, most cybersickness feature extraction methods did not place stimulus information under consideration, which predominantly influences physio-

Table 1. Cybersickness related symptoms according to 16-item SSQ [18]

No.	Symptom	Symptom Group		
		Nausea	Oculomotor	Disorientation
1	General Discomfort	✓	✓	
2	Fatigue		✓	
3	Headache		✓	
4	Eye strain		✓	
5	Difficulty focusing		✓	✓
6	Increased Salivation	✓		
7	Sweating	✓		
8	Nausea	✓		✓
9	Difficulty Concentrating	✓	✓	
10	Fullness of Head			✓
11	Blurred Vision		✓	✓
12	Dizzy (Eyes Open)			✓
13	Dizzy (Eyes Closed)			✓
14	Vertigo			✓
15	Stomach Awareness	✓		
16	Burping	✓		

logical response of VR content viewers. In addition, previous assessment works were limited in that only the resultant cybersickness level is taken into consideration without symptom level analysis. Unlike these previous works, the proposed method assesses individual cybersickness with symptom level analysis by combining VR content stimulus and physiological response, which can provide more detailed and interpretable information.

3 Proposed Method

The proposed SACA Net is divided into three parts: the stimulus symptom context guider, the physiological symptom guider, and the symptom relation embedder. Given VR content, the stimulus context guider extracts the symptom group features that represent the context of sickness-inducing stimulus environment. The physiological symptom guider utilizes physiological signals being collected from humans while watching the VR content to extract symptom features. Based on the symptom group features and the symptom features, the symptom relation embedder refines symptom features by embedding relational characteristics among symptoms to effectively estimate symptom levels. The proposed model covers 16 symptoms for cybersickness according to [18] (see Table 1).

3.1 Stimulus Symptom Context Guider

Fig. 2 network configuration of the stimulus symptom context guider for encoding symptom group features. There exists neural mismatch theory [40] that explains the process of motion sickness arising. When the expected sensory information does not match the actual sensory information, a neural mismatch

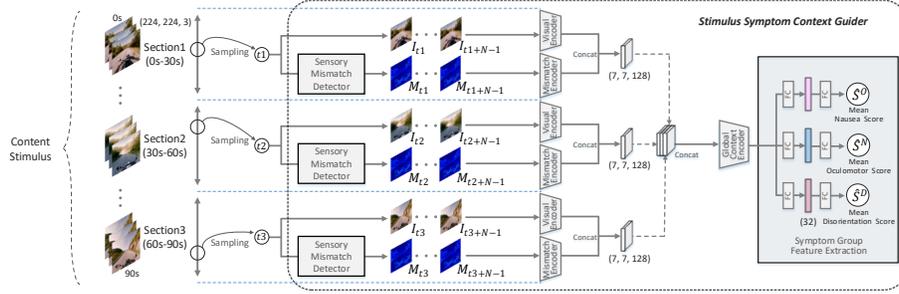


Fig. 2. Network configuration of the stimulus symptom context guider. It extracts group symptom features that represent nausea, oculomotor, and disorientation.

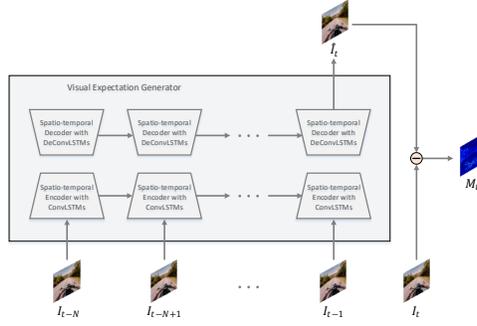


Fig. 3. Network configuration of the sensory mismatch detector in the stimulus symptom context guider. The sensory mismatch detector encodes the mismatch features.

occurs and it leads to motion sickness. People have a large neural mismatch for exceptional motions because such motions are not often experienced in daily life. Based on the theory, the sensory mismatch detector in stimulus symptom context guider is designed to encode mismatch features as shown in Fig. 3. The sensory mismatch detector includes a visual expectation generator. The visual expectation generator predicts the next frame $\hat{I}_t \in \mathbb{R}^{224 \times 224 \times 3}$ by taking previous N frames I_{t-N}, \dots, I_{t-1} ($N=11$). Note that the viewports of VR content are used as input frames. The visual expectation generator contains ConvLSTMs [50] and DeConvLSTMs [28] with deconvolution [35]. As the human daily experience, the visual expectation generator is pre-trained with videos [21] that contains only non-exceptional motions and the high frame rate (over 30Hz). Thus, frame difference is large for VR content that could induce cybersickness with exceptional motions. A pixel-wise generation loss is defined to train the generator. Let G denote the generator function. The generation loss is defined as

$$\mathcal{L}_{gen} = \|G(I_{t-N}, \dots, I_{t-1}) - I_t\|_2^2. \quad (1)$$

After training the visual expectation generator, the sensory mismatch detector takes sequence (I_{t-N}, \dots, I_t) to create mismatch feature M_t that represents

visual sensory conflict between expected and actual information. Note that the sensory mismatch detector is first pre-trained and the weights are fixed.

Based on the sensory mismatch detector, the stimulus context guider outputs symptom group features that represent nausea, oculomotor, and disorientation group factors in a context of stimulus. Given video content, three temporal sections with equal lengths are divided up. From each section, randomly sampled content video sequence (I_t, \dots, I_{t+N-1}) and mismatch sequence (M_t, \dots, M_{t+N-1}) are used as inputs at training time. Since learned combinations are diversified through random sampling, overfitting can be alleviated. Note that the midst frames of each section are sampled at testing time. Content and mismatch sequences are fed into a visual encoder and a mismatch encoder, respectively. In this process, visual context and visual mismatch of VR content for each section are encoded with 3D-Conv layers. The output features of the three sections are concatenated and fed into a global context encoder with 2D-Conv layers to consider the overall context of the content video. Finally, the fully connected layers are applied to predict the mean nausea score, mean oculomotor score, and mean disorientation score. These scores indicate the symptom group scores according to [18]. At training time, the ground truth mean score is obtained by averaging the group scores of individuals for each content. For training the stimulus context guider, symptom group score loss $\mathcal{L}_{sym}^{group}$ is defined as

$$\mathcal{L}_{sym}^{group} = \|\hat{s}_{nau} - s_{nau}\|_2^2 + \|\hat{s}_{ocu} - s_{ocu}\|_2^2 + \|\hat{s}_{dis} - s_{dis}\|_2^2 \quad (2)$$

where \hat{s}_{nau} , \hat{s}_{ocu} , and \hat{s}_{dis} are predicted symptom group scores while s_{nau} , s_{ocu} , and s_{dis} are ground truth symptom group scores. The features used for each prediction are considered as symptom group features. The symptom group features represent prior context about symptoms that can be induced by VR content stimulus. These features are utilized in the symptom relation embedder later.

3.2 Physiological Symptom Guider

The upper part of Fig. 4 shows the network configuration of the physiological symptom guider. The physiological symptom guider takes individual subject characteristics into consideration to extract symptom features. The proposed physiological symptom guider takes EEG signal acquired while watching VR content to output symptom features.

To EEG signal, a high-pass filter with 0.5Hz cut-off frequency is applied for removing baseline-drifting artifacts, and a low-pass filter with 50Hz cut-off frequency is applied for removing muscular artifacts [31]. Note that C denotes EEG channel size which corresponds with the number of acquired brain positions. After applying the frequency filters, the spectrogram image $\bar{X}_{EEG} \in \mathbb{R}^{48 \times 128 \times C}$ of the EEG signal is obtained through Short-Time Fourier Transform (STFT) [3] to consider the frequency characteristics. \bar{X}_{EEG} is fed into an EEG time-wise encoder which is composed of 1D-Conv layers. The 1D-Conv layers in the time-wise encoder are applied on the temporal axis of \bar{X}_{EEG} . Therefore, this operation does not mix the feature in frequency-wise. Based on the clinical studies [9, 25]

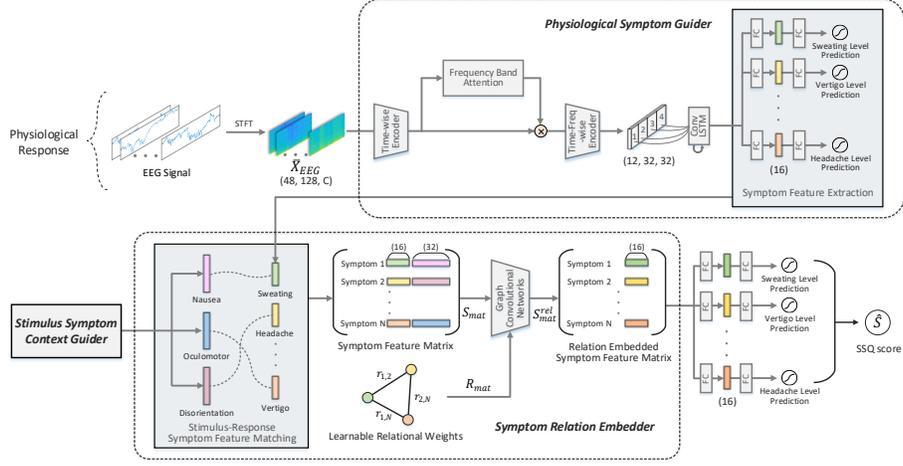


Fig. 4. Network configurations of a physiological symptom guider (upper part) and a symptom relation embedder (lower part). The physiological symptom guider outputs symptom features. The symptom relation embedder receives symptom group feature and symptom features to refine them with relational characteristics.

that show frequency bands of EEG is related to cybersickness, we design the frequency band attention encoder for emphasizing important frequency band to predict cybersickness. The frequency band attention encoder learns to obtain five attention weights that correspond with delta (0.2-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (30-50 Hz) bands. The attention weight of each band is located at the corresponding frequency region of the EEG feature map to form a frequency band attention map $A_{freq_band} \in \mathbb{R}^{48 \times 128}$. Then, A_{freq_band} is spatially elementwise multiplied to the EEG feature from the time-wise encoder. After applying the frequency band attention, the EEG feature is fed into a time-freq-wise encoder which is composed of 2D-Conv layers to encode both time and frequency characteristics. The feature drawn by the time-freq-wise encoder is divided into four patches in terms of the temporal axis. The patches enter the ConvLSTM in a temporal order. In this process, long-term characteristics can be encoded through the LSTM structure. Finally, symptom levels are predicted with fully connected layers. For training the physiological symptom guider, symptom score loss $\mathcal{L}_{sym}^{indiv}$ is defined as

$$\mathcal{L}_{sym}^{indiv} = \sum_{i=1}^{\#symptom} \|\hat{s}_{sym}^i - s_{sym}^i\|_2^2, \quad (3)$$

where \hat{s}_{sym}^i indicates predicted i -th symptom score while s_{sym}^i indicates i -th ground truth symptom score. The features used for each symptom prediction are considered as symptom features. The symptom features reflect physiological symptom characteristics related to cybersickness of individuals.

3.3 Symptom Relation Embedder

Overall procedure of the symptom relation embedder is shown in Fig. 4 . The symptom relation embedder is devised to refine the symptom features by encoding relational characteristics among symptoms. The symptom relation embedder receives the symptom group features and the symptom features from the stimulus symptom context guider and the physiological symptom guider, respectively. It learns the relations among symptoms in an unsupervised way by considering the symptom characteristics and the stimulus context that causes physical symptoms. Since symptom features are complementarily refined through embedding relations, they can be used to predict physical symptom levels more effectively.

We exploit the graph formulation [8, 27, 29, 39, 49] to learn relational characteristics. Each symptom feature $\in \mathbb{R}^{16}$ is matched to each symptom group feature $\in \mathbb{R}^{32}$. Through the symptom group matching, symptoms obtain prior context information about the sickness-inducing environment. Note that symptom features belonging to the two groups are matched to the average feature of the two symptom group features. Matched symptom features and symptom group features are concatenated and stacked row by row in a matrix form as shown in Fig. 4. As a result, a symptom feature matrix $S_{mat} \in \mathbb{R}^{16 \times 48}$ is constructed by considering each matched symptom feature as a graph node. In addition, we construct a learnable relational matrix $R_{mat} \in \mathbb{R}^{16 \times 16}$ corresponding to the adjacency matrix which represents the relationship among graph nodes. Since we set the weights of the relational matrix to be learnable, relations among symptoms can be embedded in an unsupervised way. Relational information is encoded with two layers of graph convolutional neural networks (GCNs) [27]. By following normalization trick in [27], relation embedded symptom feature matrix S_{mat}^{rel} can be formulated as

$$\tilde{R}_{mat} = R_{mat} + I, \quad (4)$$

$$S'_{mat} = ReLu(\tilde{D}^{-\frac{1}{2}} \tilde{R}_{mat} \tilde{D}^{-\frac{1}{2}} S_{mat} W_1), \quad (5)$$

$$S_{mat}^{rel} = ReLu(\tilde{D}^{-\frac{1}{2}} \tilde{R}_{mat} \tilde{D}^{-\frac{1}{2}} S'_{mat} W_2), \quad (6)$$

where I indicates identity matrix and \tilde{D} indicates diagonal node degree matrix of \tilde{R}_{mat} . $W_1 \in \mathbb{R}^{48 \times 16}$ and $W_2 \in \mathbb{R}^{16 \times 16}$ are weight matrices.

We separate S_{mat}^{rel} by each row to get each relation embedded symptom feature. Then symptom scores are predicted through fully connected layers. Symptom score loss $\mathcal{L}_{sym-rel}^{indiv}$ for relation embedded symptom features is defined as

$$\mathcal{L}_{sym-rel}^{indiv} = \sum_{i=1}^{\#symptom} \|\hat{s}_{sym-rel}^i - s_{sym}^i\|_2^2, \quad (7)$$

where $\hat{s}_{sym-rel}^i$ indicates predicted i -th symptom score by relation embedded symptom features while s_{sym}^i indicates i -th ground truth symptom score.

In addition, the individual SSQ score based on symptom features is estimated by relation embedded symptom features with fully connected layers. Individual SSQ score loss $\mathcal{L}_{SSQ}^{indiv}$ can be written as

$$\mathcal{L}_{SSQ}^{indiv} = \left\| \widehat{SSQ}_{indiv} - SSQ_{indiv} \right\|_2^2, \quad (8)$$

where \widehat{SSQ}_{indiv} is predicted individual SSQ score while SSQ_{indiv} is ground truth individual SSQ score. Finally, total objective loss can be defined as

$$\mathcal{L}_{total} = \mathcal{L}_{sym}^{group} + \mathcal{L}_{sym}^{indiv} + \mathcal{L}_{sym.rel}^{indiv} + \mathcal{L}_{SSQ}^{indiv}, \quad (9)$$

The hyper-parameters that control the balance among losses are all set to 1. The detailed network structure is included in the supplementary material.

4 Experiments

4.1 Datasets

To validate the proposed method, we conduct experiments on two public 360-degree video datasets for cybersickness assessment. Each dataset contains SSQ information [18] and corresponding physiological signals (EEG, ECG, and GSR). We employ EEG as a physiological factor to analyze cybersickness because EEG contains the most comprehensive information about the nervous system [25, 45].

VRSA DB-Shaking. In this dataset, there are 20 UHD 360-degree videos as content stimulus. The videos have various motion characteristics with camera shaking such as roller-coaster riding, skydiving, and boating. 15 subjects participated in the subjective experiment for viewing such content. Subjects were instructed to view each 90s video twice in a row, which corresponds to 180s viewing time. Repeating content twice is based on the guideline [1]. Subjects had time to rest 180s after viewing each content. Subjects graded the degree of cybersickness with the SSQ sheet [18] as [21, 43]. The SSQ sheet is composed to express the degree of 16 symptoms in 4 steps. To minimize cybersickness accumulation, subjects were asked to tell about the presence of remaining cybersickness before viewing the next content. Supplementary rest time was provided in addition to the 180s rest time until they respond ‘None at all’ as in [21, 36]. The motion of each subject was small and negligible while viewing the content. Subjects concentrated their gaze in the similar direction because used 360 degree-videos have movement in certain directions [10, 21]. Head mounted display, PIMAX 5k+ was used for presenting content. Physiological signals (EEG, ECG, and GSR) were acquired while the subjects watched the content. EMOTIV EPOC+ was used for the 14-channel EEG signal acquisition, and Cognionics AIM was used for the ECG/GSR signal acquisition. The EEG device has an acquisition sampling rate of 128 Hz, and other acquisition devices have a sampling rate of 500Hz. Experimental settings of the dataset followed the guideline of ITU-BT.500-13 [1] and BT.2021 [2].

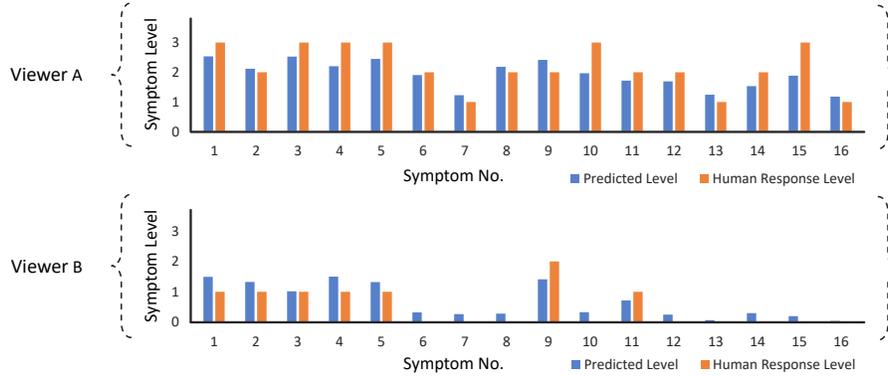


Fig. 5. Examples of symptom level prediction by the proposed method. It could differently estimate cybersickness with symptoms according to each individual viewer.

VRSA DB-FR. In addition to the VRSA DB-Shaking, VRSA DB-FR is used. This public dataset is the subject expanded version of [28]. There are 20 UHD 360-degree videos as content stimulus. The videos have two types of frame rates (10Hz, 60Hz) with various motion characteristics such as mountain biking, landscape scene, and car driving. It is known that video with exceptional motion and low frame rate causes cybersickness [21, 33, 46]. The dataset was constructed to contain various levels of cybersickness induced by content with excessive movement and low frame rate. 25 subjects participated in the subjective experiment for viewing such content. The protocol for viewing and assessment is the same as the VRSA DB-Shaking. Ultra-wide curved display, LG 34UC89 was used for presenting content. Viewing distance is controlled to provide immersive experiences with HMD level 110-degree FOV [6]. Physiological signals (EEG, ECG, and GSR) were acquired. Cognionics Quick-30 was used for 29-channel EEG signal acquisition, and Cognionics AIM was used for ECG/GSR signal acquisition. The acquisition devices have the same sampling rate of 500 Hz. Experimental settings of the dataset followed the guideline of ITU-BT.500-13 [1] and BT.2021 [2].

4.2 Implementation Details

For each content, the physiological signals are 180s long. The intermediate 120s of each physiological signal is used to remove the noise of starting and end. Data augmentation is performed by shifting the extracted 120s region by 5 seconds on the time axis. As a result, the training set is augmented 9 times. In the model training process, the stimulus symptom context guider and physiological symptom guider are first trained with their own loss functions $\mathcal{L}_{sym}^{group}$ and $\mathcal{L}_{sym}^{indiv}$ for smoothly encoding relations in the later part. In the stimulus symptom context guider and the physiological symptom guider, only the fully connected layers are learned with the final objective loss \mathcal{L}_{total} . We use Adam [26] to optimize the proposed network with a learning rate of 0.0002 and a batch size of 16.

Table 2. Symptom level prediction performances according to the network designs.

Network Design		VRSA DB-Shaking			VRSA DB-FR		
Relation Embedding	Stimulus Context	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
X	X	0.389	0.326	0.499	0.516	0.397	0.356
✓	X	0.449	0.351	0.492	0.536	0.397	0.346
✓	✓	0.478	0.385	0.441	0.574	0.427	0.328

Table 3. Detailed symptom level prediction results on the VRSA DB-Shaking.

Evaluation Metrics	Relation Embedding	Stimulus Context	Mean	# Symptom															
				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
PLCC	✓	✓	0.389	0.338	0.346	0.494	0.458	0.556	0.522	0.364	0.464	0.336	0.456	0.392	0.246	0.332	0.182	0.360	0.386
			0.449	0.414	0.402	0.492	0.542	0.568	0.542	0.330	0.532	0.350	0.480	0.552	0.278	0.386	0.254	0.496	0.568
			0.478	0.550	0.474	0.490	0.546	0.580	0.552	0.414	0.498	0.360	0.526	0.544	0.392	0.396	0.290	0.478	0.560
SROCC	✓	✓	0.326	0.322	0.338	0.462	0.318	0.434	0.380	0.246	0.386	0.330	0.364	0.396	0.234	0.338	0.226	0.162	0.292
			0.351	0.350	0.342	0.470	0.406	0.438	0.356	0.202	0.424	0.326	0.340	0.388	0.298	0.360	0.264	0.284	0.368
			0.385	0.518	0.440	0.440	0.408	0.474	0.382	0.284	0.356	0.308	0.400	0.454	0.400	0.368	0.306	0.284	0.350
RMSE	✓	✓	0.499	0.798	0.828	0.592	0.768	0.778	0.214	0.244	0.368	0.580	0.344	0.624	0.444	0.232	0.568	0.368	0.240
			0.492	0.768	0.804	0.602	0.754	0.858	0.204	0.238	0.360	0.634	0.336	0.562	0.442	0.222	0.562	0.330	0.206
			0.441	0.552	0.660	0.590	0.668	0.740	0.204	0.228	0.350	0.584	0.304	0.524	0.386	0.222	0.516	0.322	0.216

4.3 Assessment Performance Evaluation

We perform 5-fold cross-validation. The 5-fold is separated based on the content so that content and physiology in the training set and the test set do not overlap at all. Pearson linear correlation coefficient (PLCC), spearman rank order correlation coefficient (SROCC), and root mean square error (RMSE) are used as performance evaluation metrics. The PLCC and the SROCC are utilized to measure linearity and monotonicity, respectively. The RMSE metric reflects the differences between actual scores and predicted scores.

Physical Symptom Assessment. Fig. 5 shows examples of symptom level prediction by the proposed method. The proposed method could distinguish different cybersickness with distinct symptoms according to each individual viewer for each content. Table 2 shows symptom prediction performances of the proposed method with ablating network designs on both datasets. Each symptom number matches with the symptom order in Table 1. The represented performance is the average of prediction performances for all symptoms. The baseline does not utilize relation embedding and stimulus context information. The relation embedding model without stimulus context indicates the case where symptom relation embedder is applied without symptom group features from the stimulus symptom context guider. The model with symptom relation embedding predicts symptom level better than the baseline model. The final model with relation embedding and stimulus context shows better performance than the other models for all evaluation metrics on both datasets. Table 3 shows detailed

Table 4. Total SSQ score prediction performances on the VRSA DB-Shaking and the VRSA DB-FR.

VRSA DB-Shaking				VRSA DB-FR			
Method	PLCC	SROCC	RMSE	Method	PLCC	SROCC	RMSE
Skin Conductance Level Feature [19]-based Method (GSR)	0.314	0.308	43.615	Skin Conductance Level Feature [19]-based Method (GSR)	0.390	0.295	34.933
Peak Interval Feature [42]-based Method (ECG)	0.340	0.237	46.469	Peak Interval Feature [42]-based Method (ECG)	0.379	0.298	34.712
Band Power Feature [17]-based Method (EEG)	0.492	0.352	35.157	Band Power Feature [17]-based Method (EEG)	0.476	0.326	33.862
Physiological Fusion Net [28] (EEG, ECG, and GSR + Content Stimulus)	0.739	0.617	30.372	Physiological Fusion Net [28] (EEG, ECG, and GSR + Content Stimulus)	0.806	0.660	23.893
Proposed Method (EEG + Content Stimulus)	0.751	0.679	25.373	Proposed Method (EEG + Content Stimulus)	0.801	0.671	22.937

symptom level prediction results on the VRSA DB-Shaking. The final proposed model achieves the best performances. Considering that predicting the symptoms of each subject is a very challenging task, the proposed method obtains meaningful results for symptom level assessment (correlation p-value \leq 0.05). Note that the EEG acquisition device in VRSA DB-FR is the sophisticated one with more brain channels and higher sampling rates compared to VRSA-Shaking. Thus, the baseline performance of it is higher than that of VRSA DB-Shaking.

Total SSQ Assessment. The performance comparison results for total SSQ score prediction are shown in Table 4. The skin conductance level feature-based method uses features related to tonic characteristics of GSR (MSCL, SDSCL, and SKSCL) [19]. The peak interval feature-based method performs prediction using the major RR interval features of ECG (MeanRR, SDRR, pNN50, and NN50) [42]. The band power feature-based method utilizes the frequency band power of EEG [17]. The Physiological Fusion Net [28] is a deep network model that predicts individual SSQ by fusing EEG, ECG, and GSR signals with content stimulus. As shown in the table, the proposed method outperforms the existing state-of-the-art methods. The proposed method obtains correlation results for total SSQ score prediction of individuals with PLCC \geq 0.7 (p-value \leq 0.05) on the VRSA DB-Shaking and PLCC \geq 0.8 (p-value \leq 0.05) on the VRSA DB-FR. Compared to the Physiological Fusion Net [28], our model predicts cybersickness without additional use of ECG and GSR data. Furthermore, instead of just predicting the resulting total SSQ score, physical symptoms can also be predicted to interpret which symptoms constitute the cybersickness of individual viewers.

4.4 Interpretation of Relational Weights

For interpreting relations among symptom features, we visualize the relational weights in relational matrix R_{mat} . Fig. 6 shows the visualization results of relational weights. Each row and column in the matrix represent each symptom. It

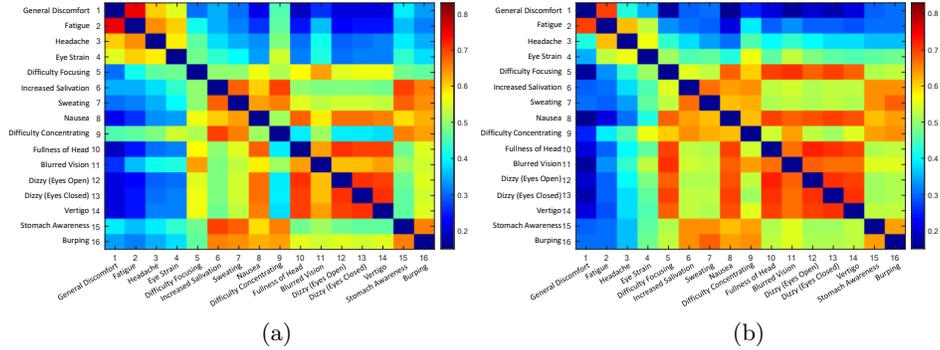


Fig. 6. Visualization results of relational weights in R_{mat} for (a) VRSA DB-Shaking and (b) VRSA DB-FR. Each row and column indicate physical symptoms.

can be seen that the symptom relational weights obtained from different datasets have a similar tendency. Looking at the most highly activated relational weights for both datasets, symptoms 10-12, 10-13 [fullness of head - dizzy] are symptoms that belong to disorientation group and are both related to head. Activated relation of 12-13 [dizzy (eyes open) - dizzy (eyes closed)] indicates closely correlated dizzy symptoms. Interestingly, the region of 1-2 [general discomfort - fatigue] is activated in weights, which are both close to the general expression of cybersickness. Besides, relational weights for each dataset contain meaningful activations of relevant symptoms such as 6-15 [increased salivation - stomach awareness, internal organ-related symptoms] in (a) and 5-11 [difficulty focusing - blurred vision, eye-related symptoms] in (b). Note that the proposed network learns relational weights in an unsupervised way. Consequently, the relational weights are convincingly learned to emphasize the relations among relevant symptoms.

5 Conclusion

In this paper, we propose the novel deep learning-based framework, SACA Net that reveals cybersickness of individual viewers with physical symptoms. Based on the physiology and stimulus context, the SACA Net effectively predicts physical symptom levels by embedding symptom relations. The symptom relation embedding scheme is designed to utilize the relational symptom characteristics for complementary effects among symptoms. The experimental results show that the proposed method achieves meaningful correlations for symptom scores and total SSQ scores on two cybersickness assessment datasets. In addition, we could interpret how the proposed SACA Net encodes relations among physical symptoms by analyzing the relational weights in the network. It is observed that relations of relevant symptoms are convincingly embedded.

Acknowledgement. This work was partly supported by IITP grant (No. 2017-0-00780), IITP grant (No. 2017-0-01779), and BK 21 Plus project.

References

1. Methodology for the subjective assessment of the quality of television pictures. ITU-R BT.500-13 (2012)
2. Subjective methods for the assessment of stereoscopic 3d tv systems. ITU-R BT.2021 (2012)
3. Allen, J.: Short term spectral analysis, synthesis, and modification by discrete fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **25**(3), 235–238 (1977)
4. Basak, D., Pal, S., Patranabis, D.C.: Support vector regression. *Neural Information Processing-Letters and Reviews* **11**(10), 203–224 (2007)
5. Bruck, S., Watters, P.A.: Estimating cybersickness of simulated motion using the simulator sickness questionnaire (ssq): A controlled study. In: *CIGV*. pp. 486–488 (2009)
6. Buck, L.E., Young, M.K., Bodenheimer, B.: A comparison of distance estimation in hmd-based virtual environments with different hmd-based conditions. *ACM Transactions on Applied Perception (TAP)* **15**(3), 1–15 (2018)
7. Carnegie, K., Rhee, T.: Reducing visual discomfort with hmds using dynamic depth of field. *IEEE computer graphics and applications* **35**(5), 34–41 (2015)
8. Chen, Y., Rohrbach, M., Yan, Z., Shuicheng, Y., Feng, J., Kalantidis, Y.: Graph-based global reasoning networks. In: *CVPR*. pp. 433–442 (2019)
9. Chuang, S.W., Chuang, C.H., Yu, Y.H., King, J.T., Lin, C.T.: Eeg alpha and gamma modulators mediate motion sickness-related spectral responses. *International journal of neural systems* **26**(02), 1650007 (2016)
10. Corbillon, X., De Simone, F., Simon, G.: 360-degree video head movement dataset. In: *Proceedings of the 8th ACM on Multimedia Systems Conference*. pp. 199–204 (2017)
11. Dennison, M.S., Wisti, A.Z., D’Zmura, M.: Use of physiological signals to predict cybersickness. *Displays* **44**, 42–52 (2016)
12. Doweck, I., Gordon, C.R., Shlitner, A., Spitzer, O., Gonen, A., Binah, O., Melamed, Y., Shupak, A.: Alterations in r-r variability associated with experimental motion sickness. *Journal of the autonomic nervous system* **67**(1-2), 31–37 (1997)
13. Egan, D., Brennan, S., Barrett, J., Qiao, Y., Timmerer, C., Murray, N.: An evaluation of heart rate and electrodermal activity as an objective qoe evaluation method for immersive virtual reality environments. In: *QoMEX*. pp. 1–6 (2016)
14. Freina, L., Ott, M.: A literature review on immersive virtual reality in education: state of the art and perspectives. In: *eLSE*. vol. 1, p. 133. " Carol I" National Defence University (2015)
15. Gallagher, A.G., Ritter, E.M., Champion, H., Higgins, G., Fried, M.P., Moses, G., Smith, C.D., Satava, R.M.: Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Annals of surgery* **241**(2), 364 (2005)
16. Grantcharov, T.P., Kristiansen, V.B., Bendix, J., Bardram, L., Rosenberg, J., Funch-Jensen, P.: Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *British journal of surgery* **91**(2), 146–150 (2004)
17. Jeong, D.K., Yoo, S., Jang, Y.: Vr sickness measurement with eeg using dnn algorithm. In: *VRST*. p. 134 (2018)
18. Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.G.: Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology* **3**(3), 203–220 (1993)

19. Kim, A.Y., Jang, E.H., Kim, S., Choi, K.W., Jeon, H.J., Yu, H.Y., Byun, S.: Automatic detection of major depressive disorder using electrodermal activity. *Scientific reports* **8**(1), 1–9 (2018)
20. Kim, H.G., Baddar, W.J., Lim, H.t., Jeong, H., Ro, Y.M.: Measurement of exceptional motion in vr video contents for vr sickness assessment using deep convolutional autoencoder. In: VRST. p. 36 (2017)
21. Kim, H.G., Lim, H.T., Lee, S., Ro, Y.M.: Vrsa net: vr sickness assessment considering exceptional motion for 360 vr video. *IEEE Transactions on Image Processing* **28**(4), 1646–1660 (2018)
22. Kim, J., Kim, W., Ahn, S., Kim, J., Lee, S.: Virtual reality sickness predictor: Analysis of visual-vestibular conflict and vr contents. In: QoMEX. pp. 1–6 (2018)
23. Kim, J., Kim, W., Oh, H., Lee, S., Lee, S.: A deep cybersickness predictor based on brain signal analysis for virtual reality contents. In: ICCV. pp. 10580–10589 (2019)
24. Kim, K., Lee, S., Kim, H.G., Park, M., Ro, Y.M.: Deep objective assessment model based on spatio-temporal perception of 360-degree video for vr sickness prediction. In: ICIP. pp. 3192–3196 (2019)
25. Kim, Y.Y., Kim, H.J., Kim, E.N., Ko, H.D., Kim, H.T.: Characteristic changes in the physiological components of cybersickness. *Psychophysiology* **42**(5), 616–625 (2005)
26. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. In: ICLR (2015)
27. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: ICLR (2017)
28. Lee, S., Kim, S., Kim, H.G., Kim, M.S., Yun, S., Jeong, B., Ro, Y.M.: Physiological fusion net: Quantifying individual vr sickness with content stimulus and physiological response. In: ICIP. pp. 440–444 (2019)
29. Li, Y., Ouyang, W., Zhou, B., Wang, K., Wang, X.: Scene graph generation from objects, phrases and region captions. In: ICCV. pp. 1261–1270 (2017)
30. Lin, C.T., Chuang, S.W., Chen, Y.C., Ko, L.W., Liang, S.F., Jung, T.P.: Eeg effects of motion sickness induced in a dynamic virtual reality environment. In: EMBC. pp. 3872–3875 (2007)
31. Lin, C.T., Tsai, S.F., Ko, L.W.: Eeg-based learning system for online motion sickness level estimation in a dynamic vehicle environment. *IEEE transactions on neural networks and learning systems* **24**(10), 1689–1700 (2013)
32. Mawalid, M.A., Khoirunnisa, A.Z., Purnomo, M.H., Wibawa, A.D.: Classification of eeg signal for detecting cybersickness through time domain feature extraction using naïve bayes. In: CENIM. pp. 29–34 (2018)
33. Meehan, M., Insko, B., Whitton, M., Brooks Jr, F.P.: Physiological measures of presence in stressful virtual environments. In: TOG. pp. 645–652 (2002)
34. Naqvi, S.A.A., Badruddin, N., Malik, A.S., Hazabbah, W., Abdullah, B.: Does 3d produce more symptoms of visually induced motion sickness? In: EMBC. pp. 6405–6408 (2013)
35. Noh, H., Hong, S., Han, B.: Learning deconvolution network for semantic segmentation. In: ICCV. pp. 1520–1528 (2015)
36. Padmanaban, N., Ruban, T., Sitzmann, V., Norcia, A.M., Wetzstein, G.: Towards a machine-learning approach for sickness prediction in 360 stereoscopic videos. *IEEE transactions on visualization and computer graphics* **24**(4), 1594–1603 (2018)
37. Pane, E.S., Khoirunnisaa, A.Z., Wibawa, A.D., Purnomo, M.H.: Identifying severity level of cybersickness from eeg signals using cn2 rule induction algorithm. In: ICIIBMS. vol. 3, pp. 170–176 (2018)

38. Patrao, B., Pedro, S., Menezes, P.: How to deal with motion sickness in virtual reality. *Sciences and Technologies of Interaction*, 2015 22nd pp. 40–46 (2015)
39. Qi, M., Li, W., Yang, Z., Wang, Y., Luo, J.: Attentive relational networks for mapping images to scene graphs. In: *CVPR*. pp. 3957–3966 (2019)
40. Reason, J.T.: Motion sickness adaptation: a neural mismatch model. *Journal of the Royal Society of Medicine* **71**(11), 819–829 (1978)
41. Rebenitsch, L., Owen, C.: Review on cybersickness in applications and visual displays. *Virtual Reality* **20**(2), 101–125 (2016)
42. Shaffer, F., Ginsberg, J.: An overview of heart rate variability metrics and norms. *Frontiers in public health* **5**, 258 (2017)
43. Singla, A., Fremerey, S., Robitza, W., Raake, A.: Measuring and comparing goe and simulator sickness of omnidirectional videos in different head mounted displays. In: *QoMEX*. pp. 1–6 (2017)
44. Tiiro, A.: Effect of visual realism on cybersickness in virtual reality. University of Oulu (2018)
45. Wagh, K.P., Vasanth, K.: Electroencephalograph (eeg) based emotion recognition system: A review. In: *Innovations in Electronics and Communication Engineering*, pp. 37–59. Springer (2019)
46. Weech, S., Kenny, S., Barnett-Cowan, M.: Presence and cybersickness in virtual reality are negatively related: a review. *Frontiers in psychology* **10**, 158 (2019)
47. Wei, C.S., Ko, L.W., Chuang, S.W., Jung, T.P., Lin, C.T.: Eeg-based evaluation system for motion sickness estimation. In: *NER*. pp. 100–103 (2011)
48. Wibirama, S., Nugroho, H.A., Hamamoto, K.: Depth gaze and ecg based frequency dynamics during motion sickness in stereoscopic 3d movie. *Entertainment computing* **26**, 117–127 (2018)
49. Woo, S., Kim, D., Cho, D., Kweon, I.S.: Linknet: Relational embedding for scene graph. In: *NIPS*. pp. 560–570 (2018)
50. Xingjian, S., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.c.: Convolutional lstm network: A machine learning approach for precipitation nowcasting. In: *NIPS*. pp. 802–810 (2015)