|  |  |
| --- | --- |
| Project | **HMD based 3D Content Motion Sickness Reducing Technology**  <<http://sites.ieee.org/sagroups-3079/>> |
| Title | **Deep learning-based VR sickness assessment** |
| DCN | **3079-18-0033-00-0002** |
| Date Submitted | **July 10, 2018** |
| Source(s) | **Hak Gu Kim** [hgkim0331@kaist.ac.kr](mailto:hgkim0331@kaist.ac.kr) **(KAIST), Yong Man Ro** [ymro@kaist.ac.kr](mailto:ymro@kaist.ac.kr) |
| Re: |  |
| Abstract | With the development of 360 degree camera capture system and head mounted display (HMD), HMD-based VR contents have attracted a lot of attention of customers and industries. In viewing VR contents with HMD, VR sickness could be induced due to visual-vestibular conflict. In particular, exceptional motion in VR content could be a critical factor leading to VR sickness. The exceptional motion could exacerbate mismatches between viewer’s motion and simulation motion of the content (a visual-vestibular conflict). In this document, we introduce a novel deep learning framework approach to access VR sickness in VR content. |
| Purpose | The goal of this document is to deal with a deep learning-based objective VR sickness assessment framework by measuring the exceptional motion for evaluating the overall degree of perceived VR sickness in viewing VR content with HMD. |
| Notice | This document has been prepared to assist the IEEE 802.21 Working Group. It is offered as a basis for discussion and is not binding on the contributing individual(s) or organization(s). The material in this document is subject to change in form and content after further study. The contributor(s) reserve(s) the right to add, amend or withdraw material contained herein. |
| Release | The contributor grants a free, irrevocable license to the IEEE to incorporate material contained in this contribution, and any modifications thereof, in the creation of an IEEE Standards publication; to copyright in the IEEE’s name any IEEE Standards publication even though it may include portions of this contribution; and at the IEEE’s sole discretion to permit others to reproduce in whole or in part the resulting IEEE Standards publication. The contributor also acknowledges and accepts that IEEE 802.21 may make this contribution public. |
| Patent Policy | The contributor is familiar with IEEE patent policy, as stated in Section 6 of the IEEE-SA Standards Board bylaws <[http://standards.ieee.org/guides/bylaws/sect6-7.html#6](http://standards.ieee.org/guides/bylaws/sect6-7.html%236)> and in *Understanding Patent Issues During IEEE Standards Development* <<http://standards.ieee.org/board/pat/faq.pdf>> |

1. Introduction

With the development of 360 degree camera capture system and head mounted display (HMD), HMD-based VR contents have attracted a lot of attention of customers and industries. By presenting unlimited field of view (FoV) to viewers in spherical domain, VR content (e.g., 360 degree video) can provide immersive viewing experience.

However, as interest of VR content has grown, viewing safety issue in VR viewing has increased. When viewers watch VR contents with a HMD, *VR sickness* can be induced by several factors such as visual-vestibular conflict, FoV, frame rate, resolution, *etc*. The VR sickness could lead to three major symptoms: 1) oculomotor symptoms including visual fatigue and focusing difficulty, 2) disorientation symptoms including dizziness and vertigo, and 3) nausea symptoms including sweating and burping.

Therefore, it is essential to develop a reliable objective VR sickness assessment (VRSA) that measures the overall degree of VR sickness in order to deal with the problem of viewing safety in VR content. There were several studies for evaluating VR sickness in HMD or virtual environments. However, most of the existing studies investigated the relation between physiological signals and subjective questionnaire (e.g., simulator sickness questionnaire (SSQ)). It is cumbersome and labor-intensive to measure the physiological signals such as electroencephalogram (EEG), electrogastrogram (EGG), galvanic skin response (GSR), and heart rate.

To address the issue, in this document, we introduce a novel deep learning-based objective VRSA framework for automatically predicting the VR sickness without cumbersome physiological measurements or subjective questionnaires[[1]](#footnote-1). In this document, a novel deep learning-based VRSA framework is devised to measure VR sickness by considering exceptional motion patterns of VR content. The exceptional motion in VR content is one of the most critical factors leading to excessive VR sickness since it could exacerbate mismatches between viewer’s motion and simulation motion of the content (i.e., visual-vestibular conflict). For example, when watching 360-degree rollercoaster video with a HMD, the simulation motion of VR content perceived by visual sensor has rapid velocity and various rotations while the physical motion of viewers perceived by vestibular sensor is nearly steady. Experimental result shows that exceptional motion has a high correlation with severe VR sickness.

1. Deep Learning-based VR Sickness Assessment
   1. Overview

Fig. 1 shows the proposed VRSA framework based on deep learning. Since there are no large-scale of VR video datasets including the corresponding subjective scores for VR sickness, it is difficult to directly train the relation between VR contents and the associated subjective scores using a regression model. To address the issue, a deep convolutional autoencoder for VRSA is devised to measure the exceptional motion patterns causing severe VR sickness. For this purpose, the proposed deep convolutional autoencoder is trained with videos including non-exceptional motion such as slow and moderate motion velocity (e.g., walking and normal driving) [Fig. 1 (a)]. As a result, the trained deep convolutional autoencoder reconstructs VR videos with non-exceptional motion, which do not lead to excessive VR sickness to viewers, well. On the other hand, the trained model does not well-reconstruct VR videos with exceptional motion patterns (e.g., racing), which are highly likely to induce excessive VR sickness. Therefore, by measuring the reconstruction error between original and the reconstructed videos, the characteristics of motion in a given VR video can be estimated by trained deep network. Then, based on the reconstruction error, the exceptional motion score can be calculated that is proportional to the degree of VR sickness.

|  |
| --- |
|  |
| (a) |
|  |
| (b) |
| Fig. 1. The proposed deep learning-based VRSA framework for predicting VR sickness score (a) in training stage and (b) in testing stage. |

* 1. Non-exceptional Motion Patterns Learning using Deep Convolutional Autoencoder

To learn the characteristics of non-exceptional motion, the proposed deep convolutional autoencoder is trained with only videos with non-exceptional motion patterns such as walking and normal driving (no acceleration and rapid turning). The proposed deep network consists of two main parts which are encoder and decoder. In the convolutional encoder, multi-level spatio-temporal feature is encoded from consecutive frames of training videos with non-exceptional motion patterns. Then the encoded spatio-temporal feature is fed to the decoder in order to reconstruct the input frames. Finally, the consecutive frames are reconstructed by convolutional decoder from the encoded feature.

In training stage, the proposed deep network, deep convolutional autoencoder, is trained by minimizing the reconstruction errors between input and output frames. For training, the objective function can be written as

|  |  |
| --- | --- |
| , | (1) |

where *W* and *H* are width and height of each frame for training, respectively. *N* is the size of mini batch. **I**(*i*,*j*,*t*) is *t*-th frame. *i* and *j* indicate pixel positions.  indicates the deep convolutional autoencoder with parameter *θ*.

* 1. Exceptional Motion Score for VRSA

After training the deep network, the exceptional motion is based on reconstruction errors. In testing, VR videos with non-exceptional motion patterns or with exceptional motion patterns can be input of the proposed VRSA method. With the trained proposed deep network, , the reconstruction error at *t*-th frame, *e*(*t*), can be represented as

|  |  |
| --- | --- |
| , | (2) |

where  is trained deep convolutional autoencoder network.

Finally, based on the reconstruction error, *e*(*t*), the proposed exceptional motion score at *t*-th frame, *sm*(*t*), is can be defined as

|  |  |
| --- | --- |
| . | (3) |

Consequently, low exceptional motion score means that the given VR video could not lead to excessive VR sickness because it does not have exceptional motion patterns. On the other hand, high exceptional motion score indicates that the given VR video could lead to severe VR sickness to viewers.

1. Subjective Assessment Experiments
   1. Datasets

To measure perceived VR sickness in subjective assessment experiment and evaluate the prediction performance of the objective VRSA method, publicly available three 360-degree videos collected from YouTube were used. Fig. 2 shows the captured frame in each VR video for testing. TABLE I shows a detailed description of the test dataset. For testing, the center region of each frame is cropped and used as an input viewport.

|  |  |  |
| --- | --- | --- |
|  | | |
| (a) | (b) | (c) |
| Fig. 2. VR videos used in subjective experiments. (a) VR video1 with slow motion, (b) VR video2 with moderate motion, and (c) VR video3 with exceptional motion | | |

TABLE I. Detailed description of 360-degree videos collected from YouTube

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No.** | **Name** | **YouTube ID** | **Resolution** | **Fps** | **Time stamp** |
| 1 | Driving around Lavender fields in Valensole, Provence in Mini, 360 video | JEr3-FzSgzk | 1280720 | 30 | 00:22  – 01:22 |
| 2 | Driving in Switzerland - Grimsel Pass - RealTime - 4K UHD - GoPro Hero4 Black Edition | wECZs7hewjY | 1280720 | 30 | 01:03  – 02:03 |
| 3 | Red Bull F1 360° Experience | ClAuhgFQpLo&t=14s | 1280720 | 30 | 00:08  – 01:08 |

* 1. Equipment
* Display: Oculus Rift CV1
* PC: Intel core i7-4770@3.4GHz, 32GB RAM
* GPU: NVIDIA GTX 1080TI
  1. Subjects

A total of 15 subjects (20 – 30 years of age) participated in the subjective assessment experiment. The subjects had normal or corrected-to-normal vision and a minimum stereopsis of 60 arcsec, which was assessed by the Randot stereo test.

* 1. Questionnaires

TABLE II shows 16-item simulator sickness questionnaires (SSQ), which is widely used to evaluate the overall degree of VR sickness in viewing the VR content with HMD. The 16-item SSQ consists of 16 physical symptoms, which are highly related to VR sickness, with a discrete four point grading scale for each symptom (0: None, 1: Slight, 2: Moderate, 3: Severe).

TABLE II. 16-item SSQ score sheet

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SSQ Symptoms** | **None: 0** | **Slight: 1** | **Moderate: 2** | **Severe: 3** |
| General discomfort |  |  |  |  |
| Fatigue |  |  |  |  |
| Headache |  |  |  |  |
| Eye strain |  |  |  |  |
| Difficulty focusing |  |  |  |  |
| Increased salivation |  |  |  |  |
| Sweating |  |  |  |  |
| Nausea |  |  |  |  |
| Difficulty concentrating |  |  |  |  |
| Fullness of head |  |  |  |  |
| Blurred vision |  |  |  |  |
| Dizzy (Eyes open) |  |  |  |  |
| Dizzy (Eyes closed) |  |  |  |  |
| Vertigo |  |  |  |  |
| Stomach awareness |  |  |  |  |
| Burping |  |  |  |  |

The SSQ scores of three major symptoms (i.e., Nausea, Oculomotor, and Disorientation) are calculated by summation of scores of each symptom include in their categories. A total SSQ score is finally obtained by combining the SSQ scores for three major symptoms.

* SSQ score for nausea

The SSQ score for nausea, *SSQNausea*, can be written as

|  |  |
| --- | --- |
| , | (4) |

where *J* is the number of subjects. *sjgd*, *sjis*, and *sjs* are subjective scores of *j*-th subject for general discomfort, increased salivation, and sweating symptoms, respectively. *sjn*, *sjdc*, *sjsa*, and *sjb* are subjective scores of *j*-th subject for nausea, difficulty concentrating, stomach awareness and burping, respectively.

* SSQ score for oculomotor

The SSQ score, *SSQOculo*, for oculomotor can be written as

|  |  |
| --- | --- |
| , | (5) |

where *sjf*, *sjh*, and *sjes* are subjective scores of *j*-th subject for fatigue, headache, and eye strain, respectively. *sjdf* and *sjbv* are subjective scores for difficulty focusing and blurred vision, respectively.

* SSQ score for disorientation

The SSQ score, *SSQDis*, for disorientation can be written as

|  |  |
| --- | --- |
| , | (6) |

where *sjfh*, *sjdzo*, *sjdzc*, and *sjv* are subjective scores of *j*-th subject for fullness of head, dizzy (eye open), dizzy (eye closed), and vertigo, respectively.

* Total SSQ score

Finally, total SSQ score, *SSQtotal*, is a combination of each SSQ score for three major symptoms, which can be written as

|  |  |
| --- | --- |
| . | (7) |

The total SSQ score ranging of 32 to 40 indicates the level of perceptible VR sickness. The total SSQ score above 40 indicates excessive VR sickness to human perception.

* 1. Procedure

TABLE III shows the subjective assessment experiment procedure for evaluating VR sickness. The subjective experiment is conducted with three VR videos including different motion patterns. In this experiment, subjects grade their perceived VR sickness scores with 16-item SSQ sheet. For experiencing VR viewing condition with a HMD, a week before the actual subjective experiments, we had subjects experience various VR contents with a HMD.

In actual subjective assessment experiment, every VR video is displayed for 2 minutes through Oculus Rift CV1. After watching the VR videos, subjects rated their perception of the VR sickness for each symptom in the SSQ sheet. We instructed 16-item SSQ sheet at the beginning of the experiment. Each subjective assessment experiment for each VR video is conducted on a different day for viewing safety of the participating subjects. During the subjective assessment, they could immediately stop and take a break if they felt any difficulty to continue due to severe VR sickness. The test was performed under approval from KAIST Institutional Review Board (IRB).

TABLE III. Subjective assessment experiment procedure for each session

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Procedure (Each day)** | | |
| **Measure** | Scoring pre SSQ | Displaying VR content  (1 min.2 repeats) | Scoring post SSQ |
| **Period** | 1 min. | 2 min. | 1 min. |

* 1. Subjective Assessment Result

Fig. 3 shows the subjective assessment result for three 360-degree videos. In Fig. 3, x-axis represents VR video number and y-axis represents total SSQ score. As shown in Fig. 3, the level of VR sickness subjects perceived is proportional to the motion magnitude. The subjective assessment result shows that subjects experience several physical symptoms related to VR sickness when watching VR video 2 and 3. In particular, VR video 3, which is a 360 degree racing video including acceleration and rapid turning, could cause severe VR sickness, compared to other two VR videos. The subjective assessment result will be available on the Web[[2]](#footnote-2),[[3]](#footnote-3)

|  |
| --- |
|  |
| Fig. 3. Total SSQ scores for three VR contents with different motion patterns. |

1. Performance Evaluation Results
   1. Datasets for Training and Testing

To train the characteristics of non-exceptional motion pattern, various public video datasets were used: UCSD Ped1, Ped2, Avenue, and KITTI benchmark datasets. UCSD Ped1, Ped2, and Avenue are videos taken by a person walking. KITTI benchmark datasets are city, road, and residential normal driving datasets. Since they do not contain acceleration and rapid rotation in video sequence, it is unlikely to induce excessive VR sickness caused by visual-vestibular conflict. TABLE IV shows the description of various videos used for network training.

In our experiment, first, the deep network was pre-trained with UCSD Ped1, Ped2, and Avenue datasets. The resolution of each video frame was resized to 128128 pixels for computational efficiency. Then, the pre-trained network was fine-tuned with KITTI benchmark dataset. In fine-tuning, the KITTI benchmark dataset was resized to 50% and randomly cropped regions with 128128 pixels were used. For data augmentation in temporal domain, three different temporal stride 1, 2, and 3 were employed.

TABLE IV. Detailed description of various videos for training

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Dataset** | **The number of video clips** | **Resolution** |
| 1 | UCSD Ped1 | 34 | 238158 |
| 2 | UCSD Ped2 | 16 | 360240 |
| 3 | Avenue | 16 | 640360 |
| 4 | KITTI benchmark | 61 | 1242375 |

To test the prediction performance of the proposed VRSA, the VR videos used in subjective experiment (Section 3.1) are used. With the proposed VRSA, the exceptional motion scores of each VR video are obtained. After that, the correlation between the SSQ scores and the proposed exceptional motion scores is calculated for evaluating the performance of the proposed method.

* 1. Performance of the proposed VRSA

|  |
| --- |
|  |
| Fig. 4. Reconstruction results for VR content 1 with slow motion (left) and VR content 3 with exceptional motion (right). First and second rows represent four consecutive original frames and the corresponding reconstructed frames, respectively. The last row indicates the reconstruction error map between first and second rows. All images are normalized in range of [0, 1]. |

Fig. 4 shows the original frames, corresponding reconstructed frames, and reconstruction error frames for VR video 1 (slow motion) and 3 (exceptional motion) in order to evaluate the performance of the proposed deep network for VRSA. As shown in Fig. 4, VR content 1 is well-reconstructed because it contains slow motion pattern, which is the trained motion characteristic during training. On the contrary, VR content 3 is not reconstructed well because it has exceptional motion such as acceleration and rapid rotation in racing. The exceptional motion pattern is not trained so that the trained network cannot provide a good reconstruction quality.

|  |
| --- |
|  |
| Fig. 5. Proposed exceptional motion scores at each frame for different VR contents. |

Fig. 5 shows the measured exceptional motion score at each frame for different VR videos with different motion patterns. As shown in Fig. 5, the exceptional motion score of VR video 1 is very low while the exceptional motion score for VR video 3 is the highest. The result shows that the exceptional motion score is proportional to the subjective assessment results in Section 3.6.

* 1. Prediction Performance

To evaluate the prediction performance of the VRSA method, Pearson linear correlation coefficient (PLCC) is employed that is one of the most commonly used measurements for prediction performance. In this experiment, PLCC between the exceptional motion score and total SSQ score is calculated. The proposed VRSA considering exceptional motion has a high correlation with subjective score for VR sickness, total SSQ score (PLCC is 0.92). The result means the exceptional motion is one of the most important factors affecting VR sickness in viewing VR content with a HMD.

1. Conclusions

To deal with VR sickness in viewing VR content with HMD, it is necessary to develop a reliable objective VR sickness assessment that automatically quantifies the level of VR sickness by analyzing VR content characteristics. The exceptional motion is one of the most critical factors causing excessive VR sickness since it can exaggerate the discrepancy between simulation motion perceived by visual cue and physical motion perceived by vestibular cue. However, it is difficult to specifically model the characteristics of various exceptional motion patterns for VR sickness assessment.

In this document, to address the problem, a novel deep leaning-based VRSA framework was devised by learning the characteristics of non-exceptional motion patterns, which do not cause excessive VR sickness. As a result, the trained deep network with videos containing non-exceptional motion can well-reconstruct the VR content with non-exceptional motion. On the other hand, it cannot reconstruct the VR content with exceptional motion. Based on the reconstruction quality, exceptional motion score can be obtained.

To evaluate the level of VR sickness, a total SSQ score is used in the subjective experiment. Experimental result shows that the correlation between the exceptional motion score and a total SSQ score is significantly high. It indicates that exceptional motion of the VR content has to be considered to assess the overall degree of VR sickness.

1. H. G. Kim, W. J. Baddar, H. 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. of the 23rd ACM Conf. Virtual Reality Software and Technology (VRST)*, Gothenburg, 2017. [↑](#footnote-ref-1)
2. www.ivylab.kaist.ac.kr [↑](#footnote-ref-2)
3. www.ivylabdb.kaist.ac.kr [↑](#footnote-ref-3)