Deep Learning-based Simulator Sickness Estimation from 3D Motion Supplemental File

Junhong Zhao*
CMIC,
Victoria University of Wellington

Kien T. P. Tran*
HIT Lab,
University of Canterbury

Andrew Chalmers
CMIC,
Victoria University of Wellington

Weng Khuan Hoh
CMIC,
Victoria University of Wellington

Richard Yao
Reality Labs, Meta

Arindam Dey
Reality Labs, Meta

James Wilmott
Reality Labs, Meta

James Lin
Reality Labs, Meta

Mark Billinghurst
Empathic Computing Lab,
The University of Auckland

Robert W. Lindeman
HIT Lab,
University of Canterbury

Taehyun Rhee†
CMIC,
Victoria University of Wellington

1 Visualization

We provide additional data in this supplementary file. A visualization of the overall game map is shown in Figure 1. The user study physical space setup is shown in Figure 2. The rotational velocity decomposition is shown in Figure 3. On the training dataset, we ran statistics on each dimension of the motion dataset and obtained an overall data range as shown in Table 1. A visualization comparing the VIMSSQ scores between genders is shown in Figure 4, and VR experience and gender in Figure 5. A raw visualization of all the data points is shown in Figure 7. We can see the raw data (Figure 7) where the columns of scatter points show the sickness scores being submitted, alternating from Task 1 to Task 2. We can observe that the line remains stable, especially on the first three entries of sickness scores (Task 1, Task 2, Task 1), indicating no clear influence of order between tasks.

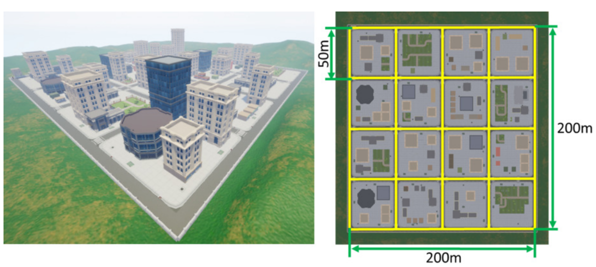

Figure 1: The game environment (left) and size 40,000m^2 (right).

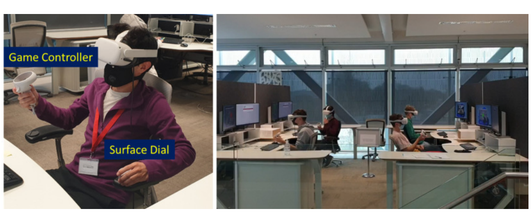

Figure 2: The physical space of the experiment with one participant (left) and with four participants in parallel (right).

*Authors contributed equally. e-mail: junhong.jennifer@gmail.com
†e-mail: taehyun.rhee@vuw.ac.nz

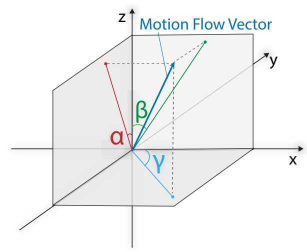

Figure 3: Schematic plot for angular calculation for roll (α), pitch (β) and yaw (γ), where $+x$ is forward, $+y$ is right and $+z$ is up.

$$\mathrm{T}_{t,(x,y,z)} = T_{t,(x,y,z)} - T_{t-1,(x,y,z)}, \tag{1}$$

$$\mathrm{R}_t = \cos^{-1}(N_t \cdot N_{t-1}), \tag{2}$$

$$\mathrm{R}_{t,yaw} = \cos^{-1}(Nx_t \cdot Nx_{t-1} + Ny_t \cdot Ny_{t-1}), \tag{3}$$

$$\mathrm{R}_{t,pitch} = \cos^{-1}(Nx_t \cdot Nx_{t-1} + Nz_t \cdot Nz_{t-1}), \tag{4}$$

$$\mathrm{R}_{t,roll} = \cos^{-1}(Ny_t \cdot Ny_{t-1} + Nz_t \cdot Nz_{t-1}), \tag{5}$$

Table 1: Data distribution of the scene features. The translation velocity and translation acceleration (relative to the camera) were divided by depth to reduce the depth effects.

Scene Features	Mean	Maximum	Minimum
Depth (cm)	17,280	65,504	15.44
Translation velocity (x)	-0.003	14.17	-14.88
Translation velocity (y)	-0.002	3.76	-6.60
Translation velocity (z)	0.000	1.02	-0.24
Translation acceleration (x)	0.006	0.50	-1.70
Translation acceleration (y)	0.010	0.81	-0.91
Translation acceleration (z)	0.004	1.68	-0.40
Rotation velocity (yaw)	0.030	0.81	-0.81
Rotation velocity (pitch)	0.010	3.07	-1.45
Rotation velocity (roll)	0.004	0.47	-1.27

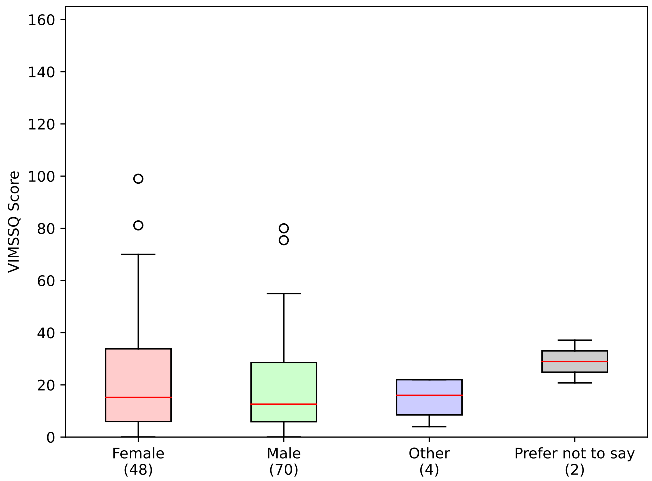

Figure 4: Comparison between each gender's VIMSSQ score.

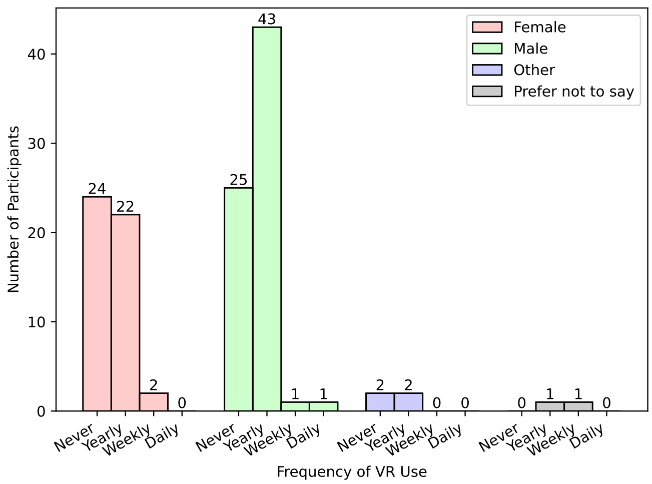

Figure 5: Comparison between each gender's VR experience.

2 Coin Collecting Task Performance

We compared the participants' performance of collected coins under the influence of the Number of Turns variable. Due to the influence of the Moving Speed factor, participants in conditions with running (RL & RH) would be able to obtain more coins than those with walking (WL & WH). Thus, we separated the test into two sets of data with two different moving speeds.

For *Walking*, a test of data normality suggested that the data were not normally distributed. As such, we used the Kruskal-Wallis H-test, which showed that there was no statistically significant difference in the number of collected coins between the two different Number of Turns settings in the walking speed configuration, $\chi^2 = 2.481$, $p = 0.115$, with a mean rank number of collected coins of 42.49 for Low Number of Turns, and 34.51 for High Number of Turns.

For *Running*, which was normally distributed, we used a paired-samples t-test, which showed that there was no significant difference in the number of collected coins between Low Number of Turns $(M = 58.61; SD = 12.99)$ and High Number of Turns $(M = 60.61; SD = 13.01); [t(37) = -0.778, p = 0.441]$.

3 Paintball Shooting Task Performance

We compared the performance factor of Correct Color Hits between the four conditions to see if there was any influence of those variables on the particular user performance. There was no statistically significant difference between conditions as determined by One-Way ANOVA $(F(3,148) = 0.416, p = 0.741)$.

In terms of Collected Coins and Correct Color Hits resulted in no clear statistical differences in investigating the influence of the Number of Turns. Even though, we may observe that the participants had slightly more correct hits when they were at a walking speed with a high Number of Turns configuration.

4 Task Performance and Simulator Sickness

We ran multiple regression tests on the performance and simulator sickness scores to observe any correlation between the two.

First, multiple regression was run to predict the simulator sickness scores for Collected Coins for the Walking conditions (WL & WH). These variables statistically significantly predicted subjective simulator sickness reported scores, $F(1,74) = 26.609, p < 0.001, R^2 = 0.264$. The Collected Coins variable added statistically significantly to the prediction, $p < 0.001$.

A second multiple regression was run with the two similar variables for the Running conditions (RL & RH). These variables statistically significantly predicted the subjective simulator sickness reported scores, $F(1,74) = 12.226, p < 0.001, R2 = 0.142$. The Collected Coins variable added statistically significantly to the prediction, $p < 0.001$.

For Task 2, we conducted a regression on the four conditions. Multiple regression was run to predict the subjective simulator sickness scores from the number of Correct Color Hits. These variables statistically significantly predicted the subjective simulator sickness scores, $F(1,150) = 17.044, p < 0.001, R2 = 0.102$. The number of Correct Color Hits variable added statistically significantly to the prediction, $p < 0.001$.

These causality tests revealed that participants who had better performance in game tasks (Collected Coins and Correct Color Hits) tend to report a lower level of simulator sickness scores.

5 Neural Network Setup

We show the details of our neural network setup in Table 2.

Table 2: Neural network setup.

Module	Layers	Kernel Size	Stride (Spatial & Temporal)	Channel	Output (t, h, w)
3D CNN (Depth)	4	(3, 3, 3)x4	(2, 2)x4; (1, 2, 1, 1)	32	(6, 2, 2)
3D CNN (Rotation)	4	(3, 3, 3)x4	(2, 2)x4; (1, 2, 1 ,1)	64	(6, 2, 2)
3D CNN (Translation)	4	(3, 3, 3)x4	(2, 2)x4; (1, 2, 1, 1)	64	(6, 2, 2)
MLP (User Susceptibility)	2	-	-	341	64
MLP (Fusion)	3	-	-	3,904	1

6 Cross-validation results

Table 3: 5-fold cross-validation results on comparisons of optical flow and motion flow for simulator sickness estimation.

Input feature	**MSE**↓	**Binary classification (Acc (%))** ↑
Optical flow	7.20	79.30
Motion Flow - Translation	5.10	81.95
Motion flow - Translation & Rotation	**4.47**	**85.87**

We also performed 5-fold cross-validation on the comparisons of optical flow and motion flow as additional results for Section 5.1. The entire dataset was divided into five equally-sized subsets. For each trial, we trained our model on four subsets while using the remaining subset for evaluation. This process was repeated five times, ensuring that each subset served as the evaluation set

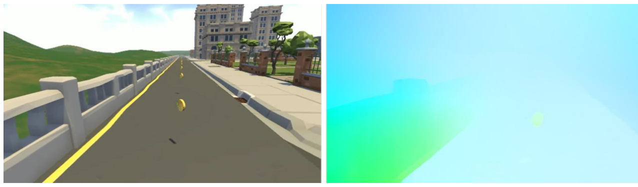

Figure 6: An example of optical flow based on the same data of Figure 8 in the main paper.

exactly once. The final results presented in Table 3 are the average performance across these five trials. We observed that the results obtained from the cross-validation align with the results obtained from testing on our dedicated testing dataset. This consistency reinforces the validity and reliability of our findings.

7 Optical Flow Example

Here we show an example of optical flow in Figure 6, based on the same data shown in Figure 6 of the main paper.

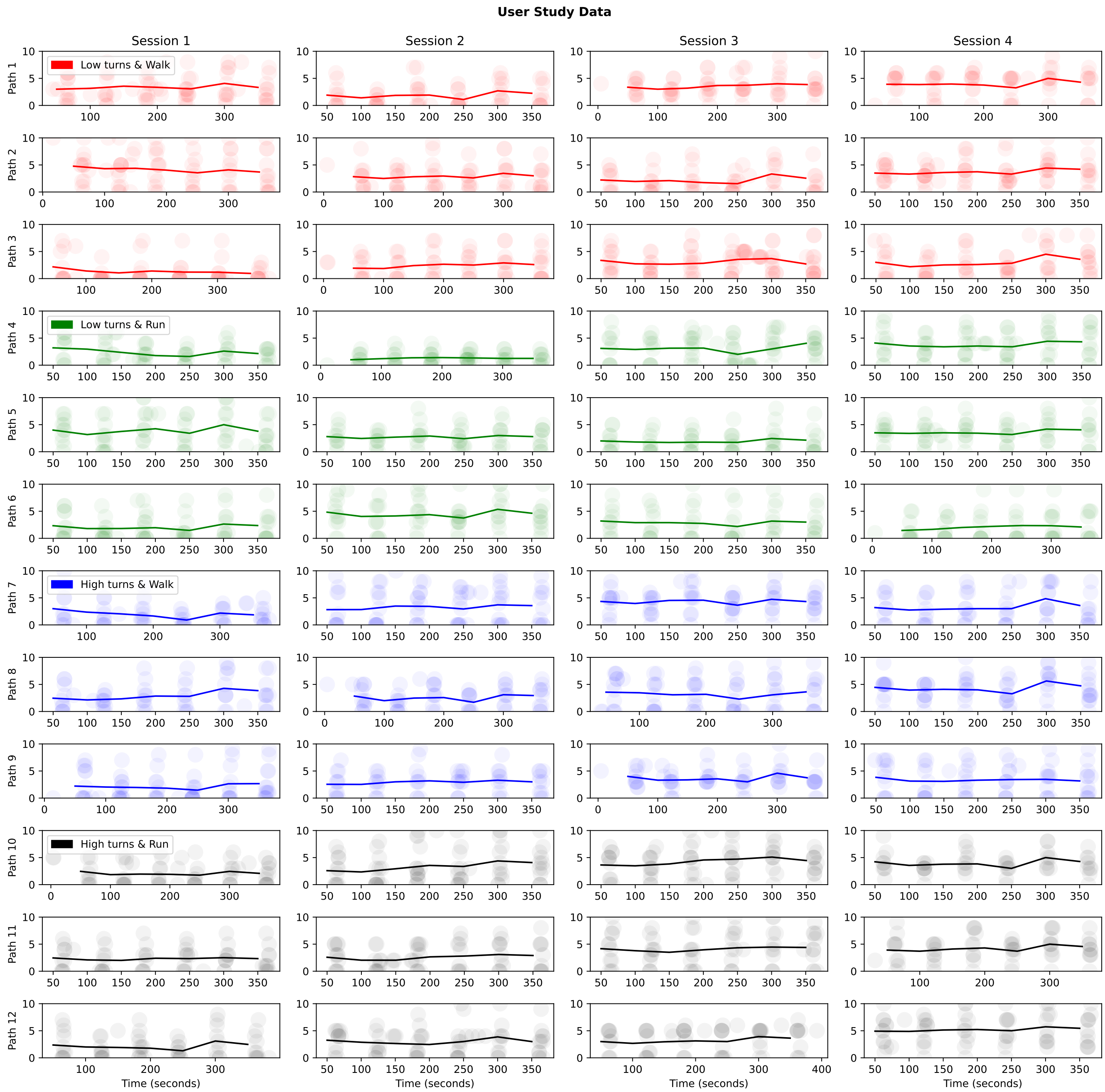

Figure 7: Visualization of the user study data. Each circle represents each simulator sickness score provided by the user. The heavily stacked columns of data are points where our system asked the participant for a score. The occasional points drifting away from the dense areas are sickness scores that were optionally submitted by the participants.