Towards a Machine-learning Approach for Sickness Prediction in 360° Stereoscopic Videos

Nitish Padmanaban*, Timon Ruban*, Vincent Sitzmann, Anthony M. Norcia, and Gordon Wetzstein

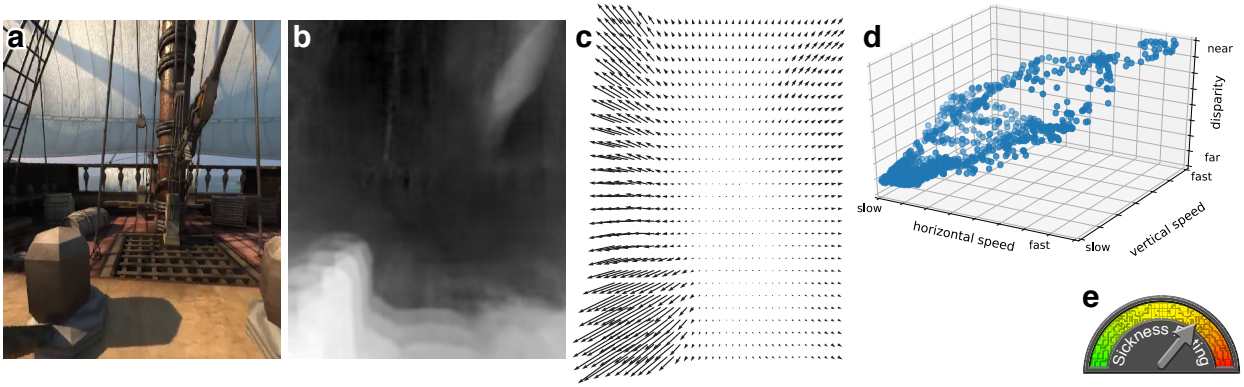

Fig. 1. We design a sickness predictor to estimate the nauseogenicity of virtual content. In designing the features for the predictor, we draw on insights from the simulator sickness literature; we also test a theory of sickness that alters the role of object depth as a feature, and we run an experiment to verify the nature of this interaction with depth. Given the result, we choose as features various summary statistics based on the interaction between depth and motion speeds in a time-varying VR video. (*a*) Shown above is the left viewport for a single frame of one of the videos in our dataset. We calculate (*b*) disparities and (*c*) optical flow vectors for each pixel as measures of the depth and motion, respectively. (*d*) We represent each pixel in a 3D vector space with axes of disparity, horizontal velocity, and vertical velocity. This representation is parameterized and binned in various ways to find the best predictor for the video's sickness rating. (*e*) The sickness rating meter displays the sickness rating in a user-friendly way.

Abstract—Virtual reality systems are widely believed to be the next major computing platform. There are, however, some barriers to adoption that must be addressed, such as that of motion sickness – which can lead to undesirable symptoms including postural instability, headaches, and nausea. Motion sickness in virtual reality occurs as a result of moving visual stimuli that cause users to perceive self-motion while they remain stationary in the real world. There are several contributing factors to both this perception of motion and the subsequent onset of sickness, including field of view, motion velocity, and stimulus depth. We verify first that differences in vection due to relative stimulus depth remain correlated with sickness. Then, we build a dataset of stereoscopic 3D videos and their corresponding sickness ratings in order to quantify their nauseogenicity, which we make available for future use. Using this dataset, we train a machine learning algorithm on hand-crafted features (quantifying speed, direction, and depth as functions of time) from each video, learning the contributions of these various features to the sickness ratings. Our predictor generally outperforms a naïve estimate, but is ultimately limited by the size of the dataset. However, our result is promising and opens the door to future work with more extensive datasets. This and further advances in this space have the potential to alleviate developer and end user concerns about motion sickness in the increasingly commonplace virtual world.

Index Terms—Virtual reality, simulator sickness, vection, machine learning

1 Introduction

As immersive virtual reality (VR) systems have become available to the average consumer, it has become increasingly clear these systems have transformative potential for how we interact with digital content, with diverse applications ranging from telesurgery to basic vision research. However, exposing users to a visual motion stimulus while they are stationary leads to motion sickness in many users [19]; as more people experience VR with the growth of the technology, the incidence of visually induced motion sickness (VIMS) may increase sharply. Furthermore, having a large field of view, which plays a role in the increased sense of presence afforded by VR, also makes users more likely to experience sickness [12]. As VR headset manufacturers pursue higher immersion, these factors make the problem of sickness likely to get worse, not better, as time passes. As of today, no complete solution exists that can effectively eliminate sickness in virtual environments without also removing desirable elements such as vection. In the absence of such a silver bullet, it becomes useful to at least quantify the amount of sickness that may occur given a 3D video. We propose methods of approaching the problem of prediction in a principled manner, drawing on techniques from machine learning to improve our performance.

In order to design a dataset and features to feed into our machine learning algorithm, we first consider the physiological causes of sickness. In particular, VIMS is caused by a visual perception of motion

- *Nitish Padmanaban*, Timon Ruban*, Vincent Sitzmann, and Gordon Wetzstein are with the Stanford Electrical Engineering Department. E-mail: gordon.wetzstein@stanford.edu.*
- *Anthony M. Norcia is with the Stanford Psychology Department.*

**Equal contribution*

Manuscript received 11 Sept. 2018; accepted 8 Jan. 2018. Date of Publication 19 Jan. 2018; date of current version 1 Jan. 2018. Author accepted version. Digital Object Identifier: 10.1109/TVCG.2018.2793560

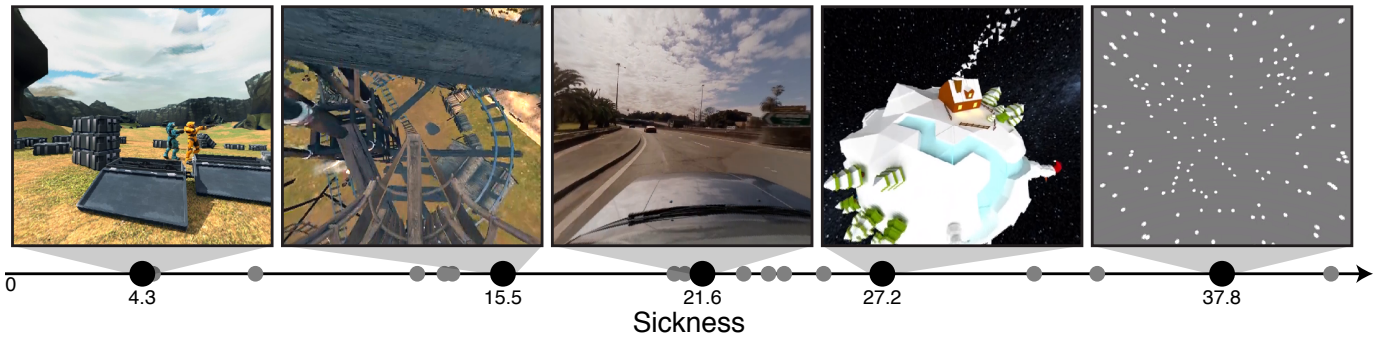

Fig. 2. We selected a set of 19 videos with widely varying scene content to create the dataset; next, we asked 96 users to each watch a single video from this set and answer the Kennedy Simulator Sickness Questionnaire. The Kennedy SSQ scores provide the basis for our ground truth sickness ratings, which are shown above (dots). A selection of videos across the spectrum of sickness have been highlighted (larger black dots).

when the user is in fact stationary, and can persist for a time even after the user has left the simulation [19]; a key theory explaining VIMS is the sensory conflict theory, which posits that sickness arises from conflicting reports from different sensory inputs, such as the visual and vestibular systems, to the brain [33]. In addition, the sickness is worse precisely *because* of the addition of stereoscopic displays [18], indicating that this problem may get worse as the display technologies are improved. Furthermore, over half of healthy adults reported experiencing simulator sickness in 3D movies [38], which provide a smaller field of view (also related to the level of immersion) than near-eye virtual displays available to consumers.

These feelings of nausea and discomfort brought about by simulator sickness pose a significant unsolved problem for any content in VR that involves movement, and imposes a limit on the potential of VR systems. For example, the Oculus guidelines instruct creators to minimize acceleration or allow reduction of FOV to reduce the possibility of sickness [28]. Many of these software-only methods for reducing sickness involve reducing field of view (FOV) [15] or vection, which may also reduce immersion [12, 19]. In choosing the optimal value for this tradeoff, one must consider that people are variable in their susceptibility to VIMS. As such, what is optimal for a moderately sensitive individual may be highly nauseogenic to another and simply lack some immersion for a third. Therefore, until some approach exists that can successfully decouple reduction of sickness from other factors that also reduce immersion – though counterintuitive, perhaps an extra-wide FOV approach [40] – it would be useful for users to have some standardized metric by which they can judge whether or not a video will induce in them an unacceptable degree of VIMS.

One way of generating this standardized metric would be to use a standardized questionnaire to rate the sickness felt after a video, for multiple users, for every video. However, as the demand and supply for immersive and exciting VR content grows however, such a method quickly becomes infeasible. We propose instead to follow this procedure for a varied dataset of stereoscopic 3D videos, and use the ratings on these videos as the basis for algorithmically rating future videos. Complicating this is the fact that, while there are models that attempt to explain how features of a scene contribute to VIMS, it is unclear how much importance should be assigned a given feature. This then provides a good fit for a machine learning approach, in which we statistically determine these relative weightings on hand-crafted features chosen based on the literature. The resulting weights of the learned model can be used to predict future videos' sickness efficiently at scale, and may also guide understanding of how differing inputs to the visual system affect VIMS.

The effectiveness of this approach is inherently dependent on the particular choice of features, but to reduce overfitting on noise in the data, we must also limit the number of features. We use the literature and our own tests as guidance on choosing the features over which to train the predictor. In particular, while velocity and FOV have been linked directly to sickness [1, 12], the effect of depth had been unclear due to the possibility of sickness-reducing reference frames [31]. Therefore, we conduct a study to confirm that relative depth of motion modulates sickness in line with vection. We conclude that velocity and depth of motion are the most useful features to include, while using a fixed FOV.

While data seems to suggest that eliminating sickness in virtual environments without hardware additions (such as galvanic vestibular stimulation and motion platforms) may not be feasible, we aim to quantify sickness to make it easier for content developers and users to predict reactions to video content. Having an objective and automated quantification process is also an important step towards further understanding of the tradeoffs involved in future VR experience design [21]. Our contributions are, in summary, that we:

- Determine that vection and sickness are correlated as a function of relative motion depth to inform the feature selection process;
- Construct a dataset of stereoscopic 3D videos over which comparative analyses of sickness may be performed;
- Build an experimental model for nauseogenicity of 3D video content using a machine-learning approach.

2 Related Work

2.1 Vection and Visually Induced Motion Sickness

When a person observes a scene, motion of objects in the scene and their own self-motion both lead to differing patterns of optical flow on the retina. This optical flow therefore contains information about the structure of the scene, but also about the direction of the person's self-motion [16]. Vection occurs when the information contained in the optical flow leads a person to perceive self-motion. On the other hand, physically induced motion sickness is a phenomenon familiar to many that experience carsickness or seasickness. When users are in a virtual environment, they experience an analog of physically induced motion sickness known as visually induced motion sickness (VIMS) that exhibits the same symptoms as physically induced motion sickness such as nausea, but also other symptoms such as blurred vision and headaches [19]. Furthermore, vection and VIMS are usually correlated with each other to the point that vection may be a prerequisite for VIMS, though some experiments do suggest that vection can be experienced without VIMS [23].

Various factors affecting vection and VIMS have been investigated; for example, the effects of multi-axis motion [4], different motion trajectories [39] (e.g. linear motion vs rotation in yaw) and velocities [1, 37], and the field of view [12, 23]. Eccentricity is another facet of the visual stimulus, but it is less clear whether it affects vection. Though earlier studies suggested that the periphery had a stronger effect on vection [3, 6], newer studies suggest that the center and periphery are equal in contribution as long as other factors such as stimulus area are carefully controlled [26].

In this work, we wish to quantify nauseogenicity of videos, which primarily provide a visual stimulus to the user; since visual stimuli

and the sickness experienced as a result are mediated by the vection response of the user, it is important to consider the factors leading to vection in constructing a varied dataset of videos and a good set of features. As such, we ensured that our dataset was varied in speeds and motion trajectories, and that the features we chose consider the magnitude optical flow in multiple directions. Since eccentricity's effect has been called into question, we do not include it among our features to limit the potential for overfitting. We also chose to not vary the FOV, since this is not intrinsic to the video content itself, but rather a function of the hardware.

2.1.1 Theories on Visual Information and Sickness

The sensory conflict theory is one of the main theories explaining the incidence of VIMS in virtual environments [33]. According to the sensory conflict theory, VIMS is caused by inconsistencies between the various streams of information that the brain receives. In this case, the visual information that facilitates vection comes into conflict with other sensory inputs which expose the fact that the user is, in reality, not moving. A similar theory states that the conflict originates not necessarily from conflicts of the sensory inputs themselves, but rather between the state of the user suggested by the sensory inputs and an internal model's expected state [5]. This also helps explain the effect of reduced sickness seen in users due to habituation [23].

Another related theory of interest is that of conflicting "rest frames," which states that the brain selects a reference frame against which to compute movement, and conflicts between the rest frames implied by the visual and other systems is what leads to VIMS [31]. Other theories to explain VIMS include the postural instability theory and the eye-movement theory, which state that the causes are changes in postural stability and optokinetic nystagmus (OKN), respectively [23]. However, these last two theories are less useful in guiding the feature selection process for our predictor, since they focus primarily on physiological phenomena as opposed to differences in video content. On the other hand, the rest frame theory has interesting consequences when considering the effect of depth ordering on vection, as discussed in the following section.

2.1.2 Effect of Stereoscopic Depth on Vection

Relative stereoscopic object depth is another important factor when considering how stereoscopic content may affect vection. The aforementioned factors affecting vection (e.g. velocity, trajectory, FOV) differ from depth in that, while they should be consistent between the stereoscopic views, they do not depend on the stereoscopic presentation itself: a "flat" display surrounding the user would still have and be influenced by these vection stimuli. On the other hand, relative object depth becomes important particularly when viewing a 3D display. It has been shown that relative object depth has a strong effect on the perception of vection. The background dominates the perception of motion direction when in motion, and strongly suppresses vection when static, regardless of fixation depth or distance separating the foreground from the background [7,20,25,29]. The foreground motion also has an effect, in that it can enhance vection when it is static or slightly counter to the background motion [27]. This foreground–background interaction can even be present when ground-truth stereoscopic cues are lacking, with whatever is perceived as background (based on other background-like qualities such as larger size or peripheral eccentricity) dominating the perception of motion [29, 34]. This perceived background is a possible reason why, in the absence of stereoscopic depth ordering, early studies saw more vection from peripheral eccentricities [26].

Considering again the rest frame hypothesis for sickness, these rest frames have been primarily used in the context of an independent visual background [14, 24, 32] that provides a static background element to reduce sickness. However, the original conception of rest frames by Prothero [31] allows for reduction of sickness with the rest frame placed either in the foreground or background. Since foreground static objects do not suppress vection as background objects do, this presents a possible venue for sickness reduction without the associated reduction in vection. Therefore, we investigate whether sickness may be reduced by using a foreground static object as the reference frame (while maintaining vection), or whether vection and sickness remain correlated as a function of depth. Depending on the outcome of this first experiment, we can then parameterize our feature selection further by dividing the motion into multiple depth layers.

2.1.3 Optical Flow Algorithms for Vection Estimation

Since optical flow on the retina determines the vection experienced by a user, it makes sense to consider the optical flow over the duration of a video shown to the user. For example, Prokop et al. showed that presenting variable optical flow patterns while someone is walking leads to instability in their walking pattern, but that they adapt over time [30] – a response that bears similarities to the postural instability and habituation associated with VIMS. In our machine learning approach, we use FlowNet, a convolutional neural network–based optical flow algorithm [13], and use its outputs to calculate our features.

2.2 Direct Vestibular Stimulation

Direct vestibular stimulation can be used to provide acceleration cues to the vestibular system, reducing or removing a major source of sensory conflict. This can be accomplished with varying degrees of efficacy using galvanic (electrical) stimulation, bone-conducted sound, or air-conducted sound [11]. By providing the vestibular system with the correct cues, the resulting reduction in sensory conflict with the visual motion should reduce VIMS. While this would remove the necessity of a predictor for videos, this requires extra hardware, introduces new safety concerns, and at least in the case of sound, can be very distracting. Unless a safe and effective vestibular stimulation system becomes available, informing users and content creators of the nauseogenic potential of videos will remain an important responsibility, which we aim to do with our machine-learning based approach.

2.3 Sickness Questionnaires

When ascertaining the potential of video content for causing sickness, we use the Kennedy Simulator Sickness Questionnaire (SSQ) [22] for users to rate the amount of sickness they felt after each video, and the Motion Sickness Susceptibility Questionnaire Short-form (MSSQ-Short) [17] to choose how to weight different users in combining their individual ratings into a single rating for each video.

The Kennedy SSQ asks participants to rate several symptoms correlated to sickness on a 4-point scale (none, slight, moderate, severe). From the 4-point scale, the Kennedy SSQ defines multiple subscores and a total score based on weighted addition of those ratings. Furthermore, it is specifically designed to apply to symptoms from simulators are opposed to real-world motion sickness, allowing us to use it for determining the nauseogenicity of virtual content.

On the other hand, the MSSQ-Short was designed to correlate well to how long it takes a given user to experience nausea given a motion sickness stimulus. It asks users how often they experience motion sickness in each of nine different situations (never, rarely, sometimes, frequently), both as a child and in the decade; these are then weighted and added, excluding situations never experienced in the given time frames. By using the individual's motion sickness history, this questionnaire allows us to account somewhat for the significant individual differences in motion sickness.

3 User Study: Relative Depth as a Feature

For the first experiment, we aim to verify whether the relative depth ordering affects sickness in proportion to its effect on vection, or whether a foreground reference frame interaction prevents sickness even with high vection from a moving background area. We find that our data supports the former, suggesting that it is important to consider depth ordering via its effect on vection when choosing features for machine learning.

3.1 Subjects

Twelve men and three women, ages 22–34, were recruited for the study. Informed consent was obtained for all participants and study procedures were approved by the institutional review board of the host

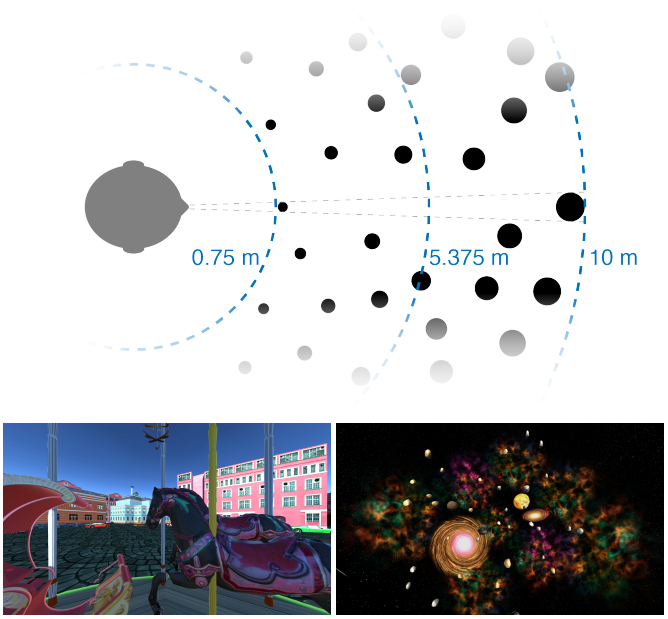

Fig. 3. (*top*) For the random dot kinematogram, the user is placed in the center of two independently rotating concentric spherical shells (as labeled) of equal width, containing dots subtending a constant visual angle. (*bottom*) The naturalistic scenes experienced by users. The near cluster of asteroids in the space scene has been brightened relative to the background for illustration.

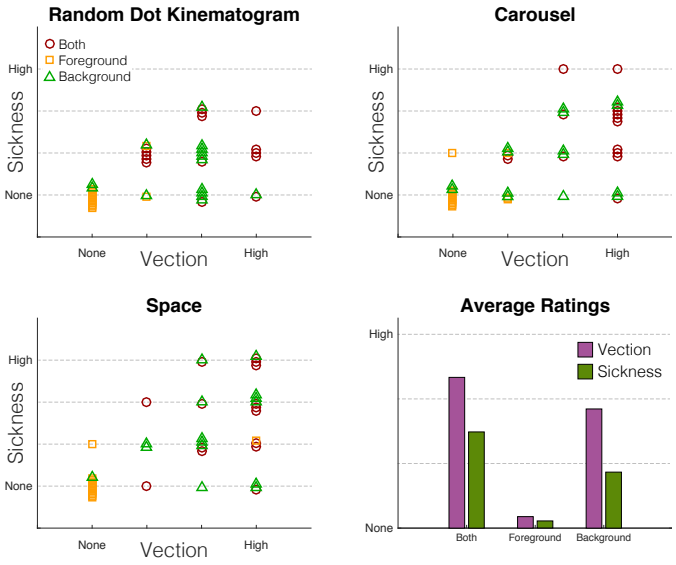

Fig. 4. The scatterplots show the vection and sickness ratings of each user in the three scenes, with motion condition indicated. The average ratings for each motion condition across all scenes is shown in the bar graph. As expected, the foreground motion condition has low ratings. The other two are similar in the relationship between vection and sickness, suggesting that nearby reference frames do not reduce sickness.

institution. All subjects had normal or corrected-to-normal vision and reported no disorders or unusual circumstances with respect to their hearing or balance, and did not report extreme susceptibility to motion sickness. Users were also screened for having stereoacuity of at least 40 arcseconds at 16″ with a Randot stereogram, corresponding to a difference in depth of 0.02″ at that distance for an average viewer.

3.2 Experiment Scenes

The test was run with three different scenes (Fig. 3), each with a foreground and background section that could rotate independently at a constant angular acceleration of $1°/s^2$ starting at 0°/s. Each scene had one of three conditions in which different combinations of the foreground and background sections are in motion. We refer to these later in the text as the foreground motion condition, the background motion condition, and the both-moving condition.

The first scene is a random dot kinematogram that surrounds the user. The dots are uniformly dispersed in depth, from 0.75–10 m, and in angular position. The extremes of the depth range are chosen to minimize the vergence–accommodation conflict by approximately staying within the zone of comfort for the 1.3 m virtual image distance of the Oculus DK2 [35]. The dots rotate about the roll axis in a randomly-chosen direction. The foreground dots are located between 0.75–5.375 m, and the background dots between 5.375–10 m so that each set of dots has an equal depth range (and therefore an equal number of dots) as shown in Fig. 3, *top*. The dots are sized such that they all subtend the same visual angle, leaving only binocular disparity as a depth cue, to ensure that the kinematogram tests primarily for stereoscopic depth ordering. The other scenes are more naturalistic (Fig. 3, *bottom*). The second scene is the carousel, in which the user is placed near the center of a carousel in the middle of a plaza. Unlike the other scenes, the user experiences rotation in yaw, with the direction chosen such that the carousel rotates in the "natural" direction of the horses relative to the surroundings. The carousel is the foreground, and the surroundings are the background. The carousel's parts are within about 1.2–5 m, and the nearest surrounding object is over 20 m away. The third scene is outer space, which is again rotation about the roll axis in a randomly chosen direction. Here, the foreground is a cluster of asteroids from 0.7–2 m, and the rest of space is the background, with the nearest object being 50 m away. Since the zone of comfort is defined in dioptric and not metric distances, these scenes also largely stay within the ideal range, despite 50 m seeming much further away.

3.3 Experiment Setup

All tests for this experiment were run using an Oculus Rift DK2 head-mounted display. This provides a resolution of 960×1080 per eye, with a 75 Hz refresh rate and a nominal 100° vertical FOV. All scenes were designed in Unity, and users were asked to only look in the forward direction, without head movement, for the duration of the trials.

When conducting the experiment, we first briefed the users on what vection and motion sickness are, then explained the reporting procedure. Each trial was displayed for 60 seconds after which the display blanked to gray. At the end of the trial, users used a controller to select the amount of vection they experienced as none, slight, moderate, or high, and the amount of sickness as none, slight, moderate, or high. While we would have preferred the Kennedy SSQ, this four-point scale was chosen to make it easier for users to rate multiple scenes without removing the headset and was sufficient for a preliminary study solely for determining utility of depth as a feature. Users that did not complete the full 60 seconds of a trial due to sickness were given the highest sickness rating. The blanked display between trials lasted for a minimum of 60 seconds as a cool-down between each trial, extending indefinitely until they were comfortable continuing. Users were free to end the study at any time.

Each user was given a total of 11 trials to rate: 2 calibration trials and 9 measurement trials. The calibration trials presented the random dot kinematogram in the foreground motion condition, then the both-moving condition, which we found in pilot testing to be most and least likely to induce vection. This was done to provide a baseline for users to calibrate their ratings. The users were not told these were baselines, and the actual random dots trials followed immediately afterwards. The other scenes were chosen to follow in a random order after the random dots scene. For each scene, all 3 motion conditions were presented together, but in a random order, to minimize drift in subjective user ratings.

3.4 Results

The vection and sickness ratings from each scene and the average rating for the motion type across scenes are presented in Fig. 4. We combine the average across all scenes since the results are largely similar. A qualitative analysis shows that the foreground motion condition elicited almost no sickness or vection, whereas the other two motion conditions are largely similar in the trend of increased vection causing increased sickness. The background motion condition is lower in both respects, likely due to the fact that the condition with both moving has twice the total optical flow. If the reference frame hypothesis were correct, we would expect to see high vection with sickness being minimal; however, we see several people that experienced the highest level of sickness with only background motion. Furthermore, the average ratings show that the ratio of the sickness to the vection rating in both conditions is very similar. To confirm this quantitatively, we assign the ratings an integer from 1 to 4 (1=None, 4=High), and take the ratio of sickness over vection for each trial. We combine motion conditions across scenes due to their similarity. This gives 1.0 for the foreground motion condition, 0.70 for background, and 0.78 for both-moving. It is fairly apparent from this ratio and the clustering in the scatterplot that the foreground motion condition is different, so we only compare the background motion and both-moving conditions for more statistical power. Despite this, the background motion and both-moving conditions fail to be significantly different ($p > 0.1$).

Therefore, we conclude that the background motion condition fails to appear different from the both-moving condition. The qualitative plots show a similar dispersal of the data in either condition, and the quantitative analysis of sickness to vection ratio are not significantly different. This indicates that depth ordering is in fact an important consideration for sickness, following the general trend of correlation between vection and sickness, and not showing any sickness reduction provided by a nearby reference frame. Based on this result, we then chose to include depth as a feature in the predictor.

4 Dataset Construction

Since we lack a well controlled dataset for videos and their sickness ratings, the first step is to construct such a dataset. We extracted 60 second clips from 16 videos from YouTube's selection of 360° stereoscopic videos; based on our previous experiment, we also included the 3 motion conditions from the random dot kinematogram to guarantee there are at least some videos capable of teasing out the depth ordering effect. This brings the total to 19 videos in our dataset. All of these videos are rendered in the omnidirectional stereo (ODS) format. We make this dataset available for future work. For our own experiments, we constrain user head motion, meaning only the forward-facing central viewport of each ODS video in our dataset was used.

4.1 Data Collection

Once the videos were selected, we gathered ratings from users. We obtained data from 96 users (76 male, 20 female, ages 19–62, mean age 28.4). All users gave informed consent according to the procedure approved by the host institution's institutional review board. Data was collected at various locations and events on the host campus using a mobile testing station.

For the data collection process, we used an HTC Vive head-mounted-display, which provides a resolution of 1080×1200 per eye, with a 90 Hz refresh rate and a nominal 110° vertical FOV. In pilot testing, when using Google Cardboard viewers to create a lower FOV dataset, we were unable to obtain appreciable measurements for sickness. As such, we chose to constrain ourselves to the HTC Vive and a single, higher FOV that elicited sickness more often. The users' head motion was constrained using a headrest to ensure that all users saw the same scene and users were asked to not look around.

Each user watched only one video during the test to minimize effects of sickness accumulation. During the trial, a user would watch the video clip for 60 seconds and then subsequently fill out the Kennedy SSQ and the MSSQ-Short. A small number of users (13) were repeated once to correct large imbalances in the number of responses per videos due to random video selection; repeated trials took place several days after the initial trial, and used a different video from what the user originally saw. This brings the total number of trials across all videos to 109.

Finally, during inspection of the survey responses for some users, we noticed that some selected "Severe" in the Kennedy SSQ items corresponding to blurry vision. When asked, they commented that the video appeared blurry from the very start. When reviewing the videos these users watched, it turned out to be a case of a low resolution video as opposed to a likely physiological effect. These ratings, while infrequent, greatly skewed the data, and as such, we excluded the SSQ questions for "blurred vision" and "difficulty focusing" from any ratings (effectively always considering them as "None").

4.2 Rating Calculation

The basis of rating calculation for each video is as follows. First, we obtain a sickness rating from each user for the video they watched. Some average all these user ratings for a given video should correspond to the intrinsic nauseogenicity of the video. However, individual users are different in their susceptibility to sickness, and therefore we also obtain a score that represents their susceptibility. We use this score to perform a weighted average of the user ratings for each video to guess that intrinsic nauseogenicity. The next several paragraphs detail this weighted averaging process.

For each trial, we calculated a single sickness rating. We start by calculating the Kennedy SSQ total score, K, as 3.74 times the sum of the nausea, oculomotor, and disorientation subscores, according to the formula given by Kennedy et al. [22]. Next, we find the motion sickness susceptibility, MS, and also the percentile P (according to the fit to percentile given by Golding et al. [17]),

$$P = 5.12\,MS - 0.0552\,MS^2 - 6.78\times10^{-4}\,MS^3 + 1.07\times10^{-5}\,MS^4. \quad (1)$$

For each video, we need to calculate a single rating, R. To do this, we need to perform a weighted average to the scores, K. We chose to do this weighting based on the MSSQ responses of each user. There are two basic ways to incorporate the susceptibility. The first is to normalize the scores so that each user is comparable by down-weighting more susceptible users relative to less susceptible users. However, there are two issues with this approach. First, many less susceptible users reported little to no symptoms, so this has the effect of making most videos seem to have low sickness ratings. Second, we are interested not in comparing users so much as we are teasing out the difference between the videos. Therefore, we use the other approach, which is to weight more susceptible users *higher*, increasing the sensitivity of our measurement for each video. This gives us as the final formula for sickness rating of video i,

$$R_i = \frac{\sum_{u\in\mathscr{U}_i}\left(\frac{wP_u}{100}+1\right)K_{i,u}}{\sum_{u\in\mathscr{U}_i}\left(\frac{wP_u}{100}+1\right)}, \quad (2)$$

where $\mathscr{U}_i$ is the set of users that watched video i, and w is free parameter that determines the relative weight between the highest and lowest percentiles. Adding 1 avoids numerical instability that may arise from values of P_u near 0. We find our data to be relatively insensitive to the choice of w, and choose $w = 2$ as a value that seems to work well.

Increasing the sensitivity in this way necessarily makes our calculation far more vulnerable to outliers; however we cannot remove outliers based on the increased-sensitivity scores, because particularly susceptible users may be detected as outliers, which would defeat the purpose of increasing sensitivity. Therefore, when checking for outliers, since we now want to make the users comparable to each other, we normalize by down-weighting the more susceptible users in accordance to their percentile this time. The normalized scores for each trial are then

$$N_{i,u} = \frac{K_{i,u}}{P_u + 1}, \quad (3)$$

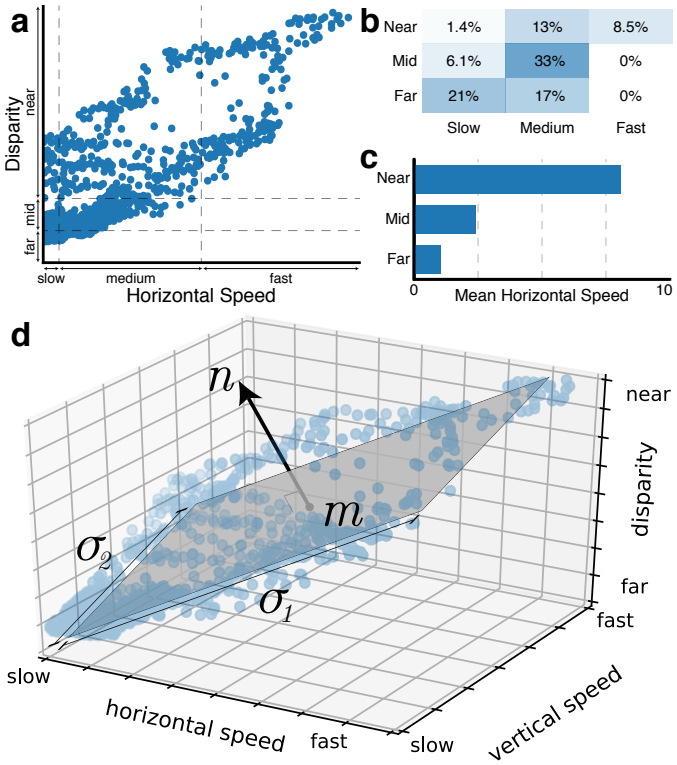

Fig. 5. An illustration of feature calculation for a single frame. (*a*) For the first two sets of features, we'll consider only a single component of the motion at a time, in this case, horizontal speed. This gives us a 2D space which the pixels occupy, which is then divided into nine regions. (*b*) The first set of features are the percent of pixels within each division of *(a)*, giving an approximation of the fraction of the visual field which that disparity–speed range represents. (*c*) The second set of features only groups the points *(a)* by disparity and uses the mean velocity in that disparity range. (*d*) The last set of features operates directly on the 3D representation of the pixels. They use PCA to extract the normal n, mean m, and explained variances σ_1^2 and σ_2^2 to capture relative interactions and spread of the data.

for video i with user u. We add 1 to avoid division errors. We found that our dataset was relatively insensitive to the exact form of this weighting as well, and as such, went with the simplest one. When we check for outliers, we consider trials across all the videos together to increase robustness of outlier detection (i.e. we remove outliers from the set of all 109 trials, as opposed to from each video's set of 5 or 6 trials), and use the 1.5 interquartile range definition for outliers. This results in us removing 8 scores as outliers. The outliers were all from different videos and these videos' final scores were also well spread over the range of sicknesses (Fig. 2). Since not all the outliers were from the same one or two videos, this suggests that the outliers really were due to individuals giving abnormally high sickness ratings as opposed to accidentally filtering out the highest sickness videos. It also suggests that testing for outliers across all videos instead of each video individually is a valid approach.

5 Predicting Sickness

The intuition gained from the literature and our own experiments point to the importance of the relative depth of objects, and the amount of movement in the virtual scene (i.e. things that affect vection strongly). Other factors, such as postural stability, attention, and optokinetic nystagmus (OKN), are not available directly from the video content itself, and require external measurements of the user. As such, we choose to focus our feature selection efforts primarily towards disparity, velocity, and their interaction.

In order to compute these features, we use optical flow algorithms. Optical flow from one frame to the next on a given viewport (left eye) can provide us with information about velocity in the scene at every pixel location, allowing us to calculate features based on eccentricity, speed, direction, and number of moving pixels in the video; however, since there are indications that eccentricity does not affect vection [26], we choose not to consider it. This also at least slightly reduces the inevitable problem of overfitting to our relatively small dataset. The optical flow algorithm was also used to compute disparity by considering the flow between the left and right viewports, due to it seeming more robust to distortion in some videos than traditional disparity matching. We also discard any vertical component of the optical flow as noise for the disparity calculation, and the horizontal component gives an offset in pixels.

Finally, before continuing, we set aside four of the 19 videos as a test set, chosen at random from within each quartile of sickness ratings, and use the other 15 videos for training.

5.1 Feature Selection

When considering features, we must balance the need for sufficient features to describe the data against the need to prevent overfitting. However, with any more than three or four features, it is almost impossible to prevent overfitting on our dataset. Therefore, we first consider a large feature space and later use forward selection to reduce the number of features to an important handful. Then, we try to see if those features match intuitions or seem more likely to have arisen from random overfitting.

All of the features were calculated on two size scales. First, on the scale of the whole video frame, and second on each quadrant of the video frame, giving five regions for each feature. This allows us to differentiate between large, slow motions and smaller, fast ones. It also accounts for types of motion such as roll that may cancel when averaged over the entire frame, but not within each quadrant. Next, to account for temporal changes, we apply a moving average for the timeseries data of each feature in each region, and take the maximum of the filtered timeseries data over the entire video,

$$f_W = \max_{t\in[W/2,60-W/2]} \frac{1}{W}\int_{t-W/2}^{t+W/2} |f(\tau)|d\tau, \tag{4}$$

where W is the length of the moving average, and $f(t)$ is the feature value (e.g. velocity or disparity) at some time t in the given region. The moving average is computed with lengths of W = 7.5, 15, 30, and 60 seconds to account for different time scales of motion. The absolute value of $f(t)$ accounts for motions with the same trajectory but opposite direction.

When selecting the actual features to compute for each spatial and temporal scale, we fall back on the intuitions from the literature and our experiments as mentioned above. Specifically, since the percept of vection is dominated by the motion of the background, we expect the velocities of objects at the furthest relative disparity to have the most effect on sickness via vection. We design several sets of features in an attempt to capture this intuition.

For the first set of features, we define three disparity bins with thresholds at 1.5 px and 4.0 px (roughly 0.15° and 0.4°), and three velocity bins with the thresholds at 1 px/s and 10 px/s (0.1°/s and 1°/s), forming a 3×3 grid for each measure of motion considered. These thresholds were chosen based on histograms of the entire dataset. The four measures of motion we use are the sum of velocity and the sum of the absolute value of velocity in each of the vertical and horizontal directions. We consider the horizontal and vertical components separately to preserve the vector nature of the motion with our scalar features. Also, considering both the sum and sum of absolute values differentiates between large camera motions (in which case the two will be similar) and movement of several objects over the scene (in which case they tend to differ due to cancellations in the sum but not in sum of absolute value). This gives four grids of binned values. Then, we count the number of pixels that fall into each of the nine bins and calculate from that the approximate percent area represented by each bin in the 2D grid. This results in 36 features per region.

Table 1. Final Chosen Features

Base feature	Spatial region	Time scale (W)
Mean disparity	Quadrant II	7.5 s
Mean vertical velocity	Quadrant IV	60 s
Medium speed, near disparity bin	Quadrant IV	15 s
Medium speed, near disparity bin	Quadrant IV	30 s
PCA n_1[1]	Quadrant IV	30 s

The second set of features uses only the three disparity bins, with the same thresholds. We still consider the horizontal and vertical velocities, but this time the velocity measures are not binned. Instead we take the average value and average of the absolute value within the given region and disparity bin. This results in another 12 features per region.

The last set of features attempts to be more data-driven in its approach. Illustrated in Fig. 5, we approximate a plane that describes the data using PCA. We consider each pixel as a point in $\mathbb{R}^3$, with the coordinates being disparity, horizontal velocity, and vertical velocity. We perform a PCA analysis with these points to get the two principal directions of highest explained variance. Since these define a plane, we can use the unit normal to the plane, n, to represent both; we constrain the first component of n to be positive for consistent results. We also include the actual explained variance in those two principal directions, σ_1^2 and σ_2^2, as another pair of features to describe the spread of the points. Finally, since PCA discards the mean, m, we add another three features for the mean of the points in each dimension, resulting in eight features per region.

5.2 Model Selection

While modern machine learning approaches often use neural network architectures to automatically select optimally relevant features in the data, they also often fundamentally require hundreds or thousands of data points to generalize well. Since our entire data set is 19 videos, we do not believe this to be the best choice of model. As such we turn to classical methods, which are less data-dependent.

Our "ground truth" measurements are obtained from real-world measurements of users. As such, we expect them to vary from some intrinsic ground truth value for each video due to noise in the measurement process. Furthermore, our features are likely have strong interdependencies due not only to the multiple temporal and spatial scales, but also to the interaction between velocity and disparity. Therefore we use bagged decision trees as our model of choice. Decision trees predict the mean of all training samples falling into distinct, non-overlapping regions, which let them capture nonlinear dependencies in the data [8]. Growing a binary decision tree for regression involves recursively branching on a random feature's value until a specified maximum depth. The resulting split at every branch minimizes the variance of the target feature as a function of the branch decision value. The maximum depth hyperparameter is adjusted to control overfitting. Combining this with bootstrap aggregating (bagging) is a proven way to reduce the variance of this method and achieves a predictor more robust to noise in the dataset [9]. Bagging entails averaging the predictions of many decision trees, each grown from a "new" dataset sampled with replacement from the original training set.

Next, since we have many more features than samples, we need to address the issue of overfitting. In order to limit the number of features – and therefore reduce overfitting – we use a beam search variant of forward selection to identify which features to use as our regressors. Forward selection is used to select a few maximally useful features from a larger feature set. This is done using a greedy algorithm that selects the best feature in isolation, the next best feature in conjunction with the first, etc. However, it tends to miss interacting features that are only useful for prediction when combined with others. On the other hand, exhaustively assessing every subset of features is exponential and intractable. Beam search is a heuristic search algorithm approximating the advantage of an exhaustive search, without the exponential runtime. At every step of the forward selection, instead of keeping only the single best feature, we keep multiple; we chose a beam width of three features per step.

Finally, to make better use of the small number of samples we have, during training we use five-fold cross-validation (CV) to choose the best model hyperparameters and perform the forward selection. This leaves us with 50 bagged trees with a maximum depth of three as the best choice of hyperparameters. The best features, as selecting by this method, are listed in Table 1. These are features that roughly account for average disparity, average vertical speed, and proportion of near-disparity motion. They also cover a range of time scales. However, when considering the spatial locations of the features, we see that most of them are in quadrant IV – this suggests that overfitting may still be an issue.

[1] This is the disparity component of the normal to the plane fit using PCA. The closer this component is to 1, the less the disparity varies relative to the spread of velocities.

Table 2. Training Set Cross-Validation Predictions

Video	Ground Truth Sickness	Predicted Sickness	Error
`sharks`	4.6	16.4	−11.8
`glowingdance`	7.8	16.4	−8.6
`dotsinner`	12.8	15.2	−2.4
`flyingcar`	13.6	15.7	−2.1
`oceancoaster`	13.9	18.3	−4.4
`woodencoaster`	15.5	18.6	−3.2
`spacevisit`	21.1	20.2	0.9
`roadcar`	21.6	16.3	5.4
`gardens`	22.9	18.7	4.2
`minecraft`	23.7	24.6	−0.9
`sculptures`	24.2	22.9	1.2
`snowplanet`	27.2	26.2	1.0
`ship`	31.9	32.3	−0.4
`dotsall`	33.9	31.5	2.4
`skyhouse`	41.1	30.2	11.0

Table 3. Test Set Predictions

Video	Ground Truth Sickness	Predicted Sickness	Error
`helicoptercrash`	4.3	17.0	−12.7
`dunerovers`	20.8	19.9	0.8
`cartooncoaster`	25.4	22.1	3.3
`dotsouter`	37.8	16.1	21.7

5.3 Results

To evaluate the performance of our model, we use as a baseline a naïve model that always predicts the mean sickness rating of the training set. The baseline achieves RMS errors of 9.7, 10.4, and 12.0 on the train set, train set with CV, and test set, respectively. In comparison, our model achieves RMS errors of 2.6, 5.3, and 12.6, respectively. The largest errors in CV occur at both the low and high extremes of the scoring range, with the predictions not being extreme enough (Table 2). At first glance, the RMS error for our model seems to fall short on the test set, predicting a relatively low score for a high-sickness video. However, when inspecting the individual scores for the videos in the test set (Table 3), we see some nuances. The error of the predictions is less than the validation RMS for two of the videos, but fails by a much wider margin on the other two. For three of the videos, the test errors are similar in nature to that of the CV errors. The one on which it diverges the most from the correct rating is the background-motion random dot kinematogram (`dotsouter`). This scene and the foreground-motion dots scene (`dotsinner` in the training set) are extremely similar across

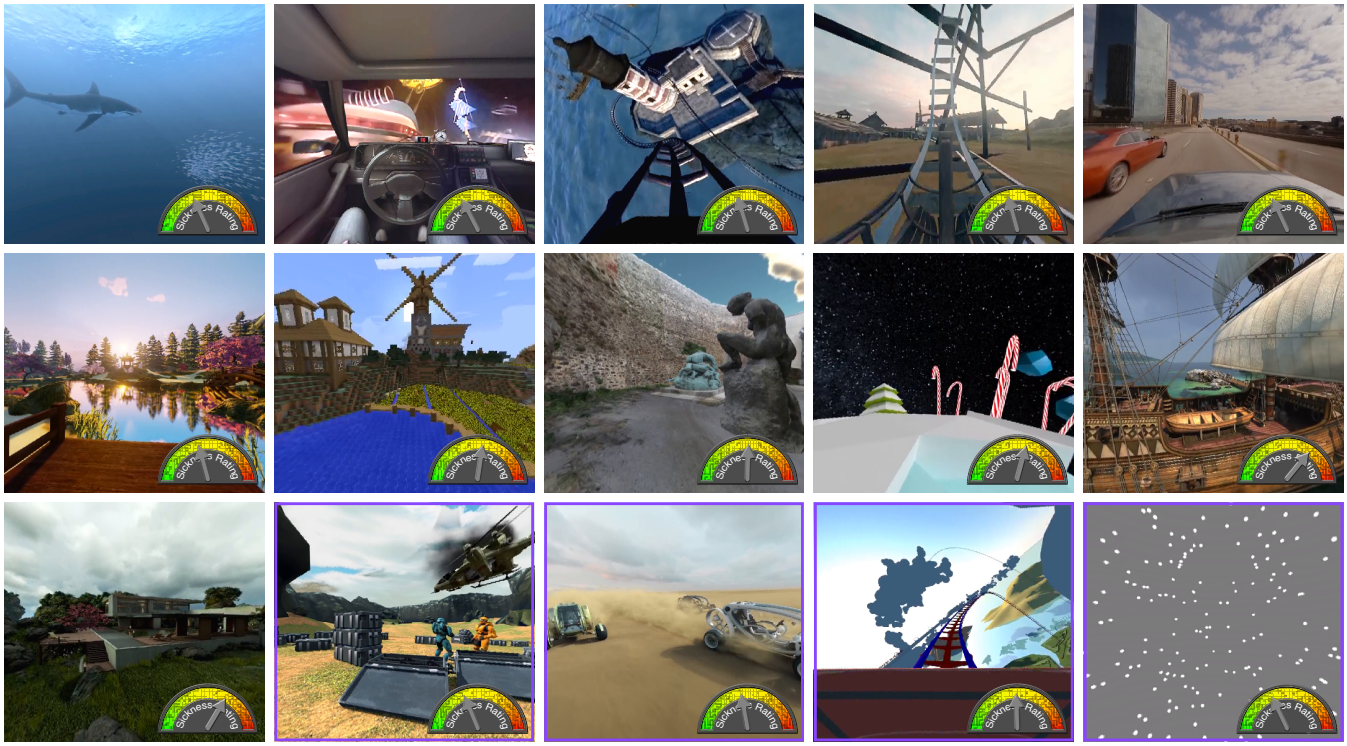

Fig. 6. The predicted sickness ratings for a selection of videos in our dataset (each point of sickness is 4° on the meter). The first 11 videos are from the training set (Table 2 in order, excluding `dotsall`, `dotsinner`, `glowingdance`, and `spacevisit`) and the last four videos (outlined in purple) are the test set (their order, left to right, corresponds to their ordering in Table 3). In general, the predicted ratings cluster near the mid-to-high range of the true sickness ratings.

most metrics, except when explicitly comparing the interaction between disparity and velocity. Since the random choice of the test and training set resulted in only one of them being in the training set, and most videos are not quite as subtle in their differences, the machine learning algorithm failed to find a differentiating feature.

6 Discussion

The results lead us to believe that the problem of predicting sickness ratings for videos using machine learning is tractable. Specifically, when we review the CV error, we see that the values are relatively close to their true values; if we consider that our true values span a range of about 40, and wish to split this into a user-facing sickness rating with four levels (e.g. none, slight, moderate, high), that gives a width of 10 per level – with the RMS of 5.3 that we achieve in CV, we should be off by at most one level on average. This low CV error indicates that the representational capacity of our features is large enough to address this challenge. Furthermore, the representational power of the features can only improve as more accurate techniques are developed for the algorithms we employ in their calculation (optical flow for example, which tends to yield noisy results at times). This, in addition to the reasons stated in the previous section, also in part explains the difficulty in correctly predicting the ratings for the background-motion kinematogram, which not only has a velocity map that is noisy, but also a disparity map that is similar to the foreground-motion kinematogram. For example, the optical flow could be used to directly calculate the user's perceived direction and speed [2] as features instead of relying on the algorithm to learn it from scratch.

On the other hand, in the test set we see issues with accurately predicting the ratings on videos. We also see that four out of five of our features after forward selection focus on just quadrant IV, and our train and CV errors differ by a factor of two. These both suggest that we are in the high variance regime, that is, that we are overfitting on noise in the data. In addition, the CV error, often a good estimate of the test error, instead differs by another factor of two from the test error. This suggests that the videos in the test and train sets are not similar enough to each other – this makes sense, we explicitly aimed to cover a wide variety of videos when we collected our examples. However, this variation turns out to be a double-edged sword, effectively forcing the machine-learning model to extrapolate a train set regressor onto a test set outside of the covered range.

Since our analysis points to the problem of assigning sickness ratings to video content being tractable, we must consider how such a rating could actually be used to help users select content. The first step involves the content source (for example, the online store or the content creators themselves), which would run the predictor on videos of the scenes or simulated gameplay in the virtual experience. Once this is done, the rating generated here should then be easily accessible or prominently displayed to any user wishing to download it. With a universal sickness rating, users could easily judge for themselves if they can handle the game or video they wish to download. Analogous to how some people avoid rollercoasters but others enjoy them, some users may wish to avoid anything with a rating of over, for example, 25, while others actively seek them out. A particularly susceptible individual may even wish to avoid anything over a much lower rating, such as 5. In this way, despite being based only on metrics derived from the videos, the sickness rating provided by our approach can be informative to a wide range of end users.

One aspect of our method that may be straightforward to generalize to a fully interactive environment is our choice to constrain head motion. Since the sensory conflict model states that it is the conflict in motion and not the presence of motion itself that causes sickness, if users were allowed to freely move their heads with the virtual scene updated accordingly and with low latency, there should be a minimal increase in sickness caused by the head movement itself. However, there is one caveat in that head rotation will change the prevailing direction of motion within the scene, and therefore greatly affect the optical flow. Since our sickness ratings were collected with users always facing forward, scenes that did not exhibit changes in the direction of motion

may no longer have accurate ground truth ratings after allowing for head rotation. Therefore, we did not implement this in our current setup, but if sickness ratings are collected for a set of salient viewpoints, these ratings and viewpoints could be combined and run through the same overall model.

6.1 Future Work

Many of the issues mentioned above have a common solution, which is the acquisition of a larger dataset, on the order of several hundred or more videos. While this is beyond the scope of our work, a predictor trained on a larger dataset generalizes to new data more easily. A predictor that generalizes well has further benefits, such as being able to investigate which features and interactions are actually important for prediction of sickness, beyond what is currently known in the literature. For example, while it certainly may be possible with our dataset, a larger dataset aids in testing new, more powerful features to confirm hypotheses about visual cues relevant to sickness.

A further advantage of using machine learning as a tool for prediction is that it becomes easy to extend in the future. With larger datasets come the possibility of using neural networks, for example. It is also easy for future work, given a standardized dataset of videos, to augment the questionnaire-based metrics with physiological indictors of sickness [10] as ground-truth sickness ratings, perhaps allowing for more real-time evaluations of nauseogenicity of small section of videos. Another important direction in which the current model can be extended is to take into account user behavior within the scenes. This could take the form of a saliency predictor [36] that informs the viewpoints and head movements which are in turn used to generate videos fed to the our sickness predictor. Another possible approach is to process the entire 360° video instead of smaller viewports; using a neural network, this would likely also learn something similar to saliency at an intermediate step, but potentially tailored to parts of the scene that are more likely to cause sickness. Finally, saliency and other models of user behavior may be key in extending our approach another step further to fully interactive virtual reality.

An interesting application for this is individualized ratings. A VR enthusiast could watch our set of videos, which could be done over the course of a few hours (allowing for breaks between videos), and rate each of them using the Kennedy SSQ. These scores would become their ground-truth ratings, without the need for any complicated weighting from the MSSQ, and be fed in directly to the machine learning model. Our current model can be trained quite quickly, which then gives the user a personalized classifier for future VR content. As the models and datasets become larger, this personalized use will become less viable, but the predictor itself will become more robust, reducing the need for individualized ratings.

7 Conclusion

Motion sickness in VR is a widespread problem that promises, without intervention, to only get worse as head-mounted display capabilities such as field of view are improved, despite simultaneous improvements to mitigating factors such as latency. Hardware solutions have potential safety issues, thus we propose a principled software-based approach for building a prediction model for video nauseogenicity. Coupled with a large dataset and other insights from future work using our approach, these models may in turn be used to understand visually induced motion sickness on a deeper level.

Acknowledgments

Special thanks to the following content creators, artists, and studios who were kind enough to contribute their VR videos to our dataset: Jake Mathew, Jean-Pascal Martin, Electric Lens Co, Pingwano, French Touch Records, Commodity Games, Rooster Teeth, VR Visio, GlobalVision Communication, Curiscope, Sehsucht, Nice Shoes, Matt Hermans, OniricFlow VR, Gone Coyote, and VR360 Playground.

References

[1] R. S. Allison, I. P. Howard, and J. E. Zacher. Effect of field size, head motion, and rotational velocity on roll vection and illusory self-tilt in a tumbling room. *Perception*, 28(3):299–306, 1999.

[2] F. Argelaguet and M. Maignant. Giant: stereoscopic-compliant multi-scale navigation in ves. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, pp. 269–277. ACM, 2016.

[3] A. Berthoz, B. Pavard, and L. Young. Perception of linear horizontal self-motion induced by peripheral vision (linearvection) basic characteristics and visual-vestibular interactions. *Experimental brain research*, 23(5):471–489, 1975.

[4] F. Bonato, A. Bubka, and S. Palmisano. Combined pitch and roll and cybersickness in a virtual environment. *Aviation, Space, and Environmental Medicine*, 80(11):941–945, 2009.

[5] J. E. Bos, W. Bles, and E. L. Groen. A theory on visually induced motion sickness. *Displays*, 29(2):47–57, 2008.

[6] T. Brandt, J. Dichgans, and E. Koenig. Differential effects of central versus peripheral vision on egocentric and exocentric motion perception. *Experimental Brain Research*, 16(5):476–491, 1973. doi: 10.1007/BF00234474

[7] T. Brandt, E. R. Wist, and J. Dichgans. Foreground and background in dynamic spatial orientation. *Attention, Perception, & Psychophysics*, 17(5):497–503, 1975.

[8] L. Breiman. *Classification and regression trees*. Chapman & Hall/CRC, 1984.

[9] L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996. doi: 10.1007/BF00058655

[10] 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. *International Journal of Human–Computer Interaction*, pp. 1–15, 2017.

[11] I. S. Curthoys, V. Vulovic, A. M. Burgess, E. D. Cornell, L. E. Mezey, H. G. MacDougall, L. Manzari, and L. A. McGarvie. The basis for using bone-conducted vibration or air-conducted sound to test otolithic function. *Annals of the New York Academy of Sciences*, 1233(1):231–241, 2011.

[12] P. DiZio and J. R. Lackner. Circumventing side effects of immersive virtual environments. In *HCI International (2)*, pp. 893–896, 1997.

[13] A. Dosovitskiy, P. Fischer, E. Ilg, P. Häusser, C. Hazırbaş, V. Golkov, P. v.d. Smagt, D. Cremers, and T. Brox. Flownet: Learning optical flow with convolutional networks. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 2758–2766, Dec 2015.

[14] H. B.-L. Duh, D. E. Parker, and T. A. Furness. An "independent visual background" reduced balance disturbance envoked by visual scene motion: implication for alleviating simulator sickness. In *Proc. SIGCHI*, pp. 85–89, 2001.

[15] A. S. Fernandes and S. K. Feiner. Combating VR sickness through subtle dynamic field-of-view modification. In *Proc. 3DUI*, pp. 201–210. IEEE, 2016.

[16] J. J. Gibson. *The Perception Of The Visual World*. Boston: Houghton Mifflin, 1950.

[17] J. F. Golding. Predicting individual differences in motion sickness susceptibility by questionnaire. *Personality and Individual differences*, 41(2):237–248, 2006.

[18] J. Häkkinen, M. Pölönen, J. Takatalo, and G. Nyman. Simulator sickness in virtual display gaming: A comparison of stereoscopic and non-stereoscopic situations. In *MobileHCI*, pp. 227–230, 2006.

[19] L. J. Hettinger and G. E. Riccio. Visually induced motion sickness in virtual environments. *Presence: Teleoper. Virtual Environ.*, 1(3), 1992.

[20] I. P. Howard and T. Heckmann. Circular vection as a function of the relative sizes, distances, and positions of two competing visual displays. *Perception*, 18(5):657–665, 1989.

[21] J. Jerald. *The VR Book: Human-Centered Design for Virtual Reality*. Association for Computing Machinery and Morgan & Claypool, New York, NY, USA, 2016.

[22] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.

[23] B. Keshavarz, B. E. Riecke, L. J. Hettinger, and J. L. Campos. Vection and visually induced motion sickness: how are they related? *Frontiers in Psychology*, 6:472, 2015.

[24] J. J.-W. Lin, H. Abi-Rached, D.-H. Kim, D. E. Parker, and T. A. Furness. A "natural" independent visual background reduced simulator sickness. *Pro-*

ceedings of the Human Factors and Ergonomics Society Annual Meeting, 46(26):2124–2128, 2002.

[25] S. Nakamura. Depth separation between foreground and background on visually induced perception of self-motion. *Perceptual and motor skills*, 102(3):871–877, 2006.

[26] S. Nakamura. Effects of stimulus eccentricity on vection reevaluated with a binocularly defined depth. *Japanese Psychological Research*, 50(2):77–86, 2008.

[27] S. Nakamura and S. Shimojo. Critical role of foreground stimuli in perceiving visually induced self-motion (vection). *Perception*, 28(7):893–902, 1999.

[28] Oculus. Introduction to best practices, 2017. Accessed May 10, 2017.

[29] M. Ohmi, I. P. Howard, and J. P. Landolt. Circular vection as a function of foreground-background relationships. *Perception*, 16(1):17–22, 1987.

[30] T. Prokop, M. Schubert, and W. Berger. Visual influence on human locomotion modulation to changes in optic flow. *Experimental brain research*, 114(1):63–70, 1997.

[31] J. D. Prothero. *The Role of Rest Frames in Vection, Presence and Motion Sickness*. PhD thesis, University of Washington, 1998.

[32] J. D. Prothero, M. H. Draper, T. Furness 3rd, D. E. Parker, and M. J. Wells. The use of an independent visual background to reduce simulator side-effects. *Aviation, Space, and Environmental Medicine*, 70(3 Pt 1):277–283, 1999.

[33] J. T. Reason and J. J. Brand. *Motion sickness.* Academic press, 1975.

[34] T. Seno, H. Ito, and S. Sunaga. The object and background hypothesis for vection. *Vision research*, 49(24):2973–2982, 2009.

[35] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision*, 11(8):11, 2011.

[36] V. Sitzmann, A. Serrano, A. Pavel, M. Agrawala, D. Gutierrez, and G. Wetzstein. Saliency in VR: how do people explore virtual environments? *CoRR*, abs/1612.04335, 2016.

[37] R. H. So, W. Lo, and A. T. Ho. Effects of navigation speed on motion sickness caused by an immersive virtual environment. *Human Factors*, 43(3):452–461, 2001.

[38] A. G. Solimini. Are there side effects to watching 3d movies? a prospective crossover observational study on visually induced motion sickness. *PLoS ONE*, 8(2):1–8, 2013.

[39] L. C. Trutoiu, B. J. Mohler, J. Schulte-Pelkum, and H. H. Bülthoff. Circular, linear, and curvilinear vection in a large-screen virtual environment with floor projection. *Computers & Graphics*, 33(1):47–58, 2009.

[40] R. Xiao and H. Benko. Augmenting the field-of-view of head-mounted displays with sparse peripheral displays. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 1221–1232. ACM, 2016.