·Article·

Detection of scene-irrelevant head movements via eye-head coordination information

Xiaoxiong FAN[1*], Yun CAI[1], Yufei YANG[1], Tianxing XU[1], Yike Li[1], Songhai ZHANG[1], Fanglue ZHANG[2]

1. *Tsinghua University, Beijing* 100084*, China*

2. *Victoria University of Wellington, Wellington, New Zealand*

*** Corresponding author,** fxx18@mails.tsinghua.edu.cn

Received: 30 May 2021 **Revised:** 31 July 2021 **Accepted:** 12 August 2021

Citation: Xiaoxiong FAN, Yun CAI, Yufei YANG, Tianxing XU, Yike Li, Songhai ZHANG, Fanglue ZHANG. Detection of scene-irrelevant head movements via eye-head coordination information. Virtual Reality & Intelligent Hardware, 2021, 3(6): 501—514

DOI: 10.1016/j.vrih.2021.08.007

Abstract **Background** Accurate motion tracking in head-mounted displays (HMDs) has been widely used in immersive VR interaction technologies. However, tracking the head motion of users at all times is not always desirable. During a session of HMD usage, users may make scene-irrelevant head rotations, such as adjusting the head position to avoid neck pain or responding to distractions from the physical world. To the best of our knowledge, this is the first study that addresses the problem of scene-irrelevant head movements. **Methods** We trained a classifier to detect scene-irrelevant motions using temporal eye-head-coordinated information sequences. To investigate the usefulness of the detection results, we propose a technique to suspend motion tracking in HMDs where scene-irrelevant motions are detected. **Results/Conclusions** Experimental results demonstrate that the scene-relevancy of movements can be detected using eye-head coordination information, and that ignoring scene-irrelevant head motions in HMDs improves user continuity without increasing sickness or breaking immersion.

Keywords Virtual reality; Human-centered computing; Human-computer interaction (HCI); Interaction paradigms; HCI design and evaluation methods; User models

1 Introduction

Virtual reality over head-mounted displays (HMDs) is known for its immersive experiences and has been applied to various fields, such as education[1] and healthcare[2]. View control in HMDs is usually guided by tracking head rotations of the user, as observed in common consumer-level products on the market, such as the Oculus Quest or the HTC Vive Pro. Therefore, several researchers have focused on obtaining accurate and sensitive head tracking in HMDs[3], as well as in the theoretical analyses regarding the turning behaviors and technical solutions for redirecting users[4].

In most consumer-level HMD applications, head movements are continuously tracked to drive the direction and position of the view in a virtual environment. However, users may perform actions unrelated to the VR environment in certain scenarios. For instance, a number of users complain of symptoms such as

neck pain and headaches[5] after long-term and continuous use of HMDs; such users can be made to perform active head movements and change their postures in order to mitigate the symptoms. Users can also be distracted by real-world events and perform subconscious actions when other people touch or talk to them while indulging in immersive VR activities. Such actions were found to hinder the user's sense of presence or present visual discomfort to the users[6–8]. Tracking these movements may significantly weaken the immersion characteristics of VR systems. Continuous reactions to head movements are not optimally suited for HMDs; thus, an accurate detection of these movements is required.

We define movements invoked by the VR content or aim to interact with the VR content as scene-relevant movements, while defining movements caused by other factors outside of the VR content as scene-irrelevant movements. To the best of our knowledge, no existing studies have thus far been conducted regarding the identification of scene-irrelevant movements and the improvement of VR experiences with the consideration of these movements. Our study is the first to investigate the detection of scene-irrelevant movements and our results can be used to improve the overall VR experience by reducing the impact of these movements once they are detected.

Inspired from previous studies[9,10] on gaze and head movements, we propose a novel method to identify scene-irrelevant head rotations based on eye-head coordination behavior. The proposed method was trained and evaluated on a self-collected dataset, and an acceptable accuracy was achieved. To verify the applicability of the identification results, we designed a "view following technique", which can be triggered by detected scene-irrelevant movements to automatically rotate the virtual view to follow the user head rotation. Well-designed user studies were conducted to utilize the technique in practical situations. The experimental results demonstrate that user discomfort can be significantly mitigated using the view-following technique, which in turn indicates the usefulness of the identification process.

The following three contributions were made in this study:

• The detection of scene-irrelevant movements was performed for the first time.

• A novel method was presented for identification based on eye-head coordination behavior.

• The applicability of the identification results was verified in a real-life scenario by designing a view-following technique and performing a user study.

2 Related work

2.1 Negative feedbacks in HMD view control

Existing view-control techniques highly depend on head rotations; however, this can lead to prominent stress on the head and neck and make real-world distractions more likely to cause apparent interruptions in VR experiences. Wille et al. found a higher strain and slower performance while working with an HMD compared to a tablet PC, as well as a stronger increase in the strain over longer usage times[5]. Oh et al. found that real-world distractions had a negative effect on recognition, recall, and social presence during a virtual experience, stating that more research is required regarding whether and how real-world events should be integrated into virtual environments[8]. Owing to muscle stress, relaxation actions occur almost inevitably while using HMDs; furthermore, real-world distractions are also unavoidable and unpredictable. However, to the best of our knowledge, no existing work has been conducted to reduce the impacts of head movements caused by relaxation actions or real-world distractions, a gap that we aim to fill with our study.

2.2 Gaze behavior in exploration of VR environment

Numerous studies have been conducted to improve the user experience of HMD based on a given behavior.

Gandrud et al. combined eye gaze data, head position, and head orientation information to infer the prospective heading direction of a user after reaching a distant decision point[4]. Zank et al. demonstrated that a user's eventual turn direction can be correctly predicted based on the gaze direction sooner than the positional data[11]. In recent years, an increasing number of studies have focused on the understanding of gaze behavior in VR environments; for example, Freedman et al. conducted a detailed analysis of the eye and gaze statistics and indicated that the head follows a gaze with an average delay of 58ms[12]. The aforementioned studies inspired us to utilize eye movement data to analyze the intentions of user movements tracked in HMDs.

2.3 Eye-head coordinated movements

Coordinated movements between the head and eye have been studied in cognitive psychology and neuropsychology[13,14] for decades. In particular, several studies[10,15–17] have focused on varying eye-head coordination patterns under different conditions, analyzing when and why a certain pattern appears, and verified and measured the latency between the eye and head in each pattern. Furthermore, a few studies[10,18] have proven that eye-head coordinations that are similar to the physical world also exist in VR environments. From the psychological perspective, previous studies have provided a solid theoretical basis for our research, based on which we propose using the coordination between eye and head movements to identify whether actions are relevant to the virtual scene.

In addition, recent studies have focused on exploring the use of eye-head coordinated movements in possible real-life scenarios. Doshi et al. explored eye-head movement patterns under different attention-shift conditions[9]. Their findings demonstrate that it is possible to apply measurements of eye-head dynamics for detecting driver distractions, as well as for classifying the human attentive status in time and safety critical tasks. This is similar to detecting scene-irrelevant movements, which further inspired our research.

3 Scene-irrelevant movement detection

Our study focuses on reducing the negative impact of scene-irrelevant movements on the immersion characteristics of a VR system based on the eye-head coordination behavior. We first define scene-the relevant and scene-irrelevant movements and categorize them into several types. We then conduct pilot experiments to verify the correlation between the eye-head coordinated movements and scene-relevance. Based on previous observations, we propose a novel method that uses eye-head coordinate data as the input and distinguishes between scene-relevant and irrelevant movements.

3.1 Scene-relevant and scene-irrelevant movements

We define scene-relevant movements as those invoked by, or aiming to interact with, the VR content. Correspondingly, scene-irrelevant movements are defined as movements caused by factors other than the VR content. Based on the observations of day-to-day behaviors of gazing and head motions while using VR HMDs, we propose categorizing the head movements, which provides a guidance for the design of the pilot experiment, as follows:

(1) Authors classified the scene-relevant head rotations into the following three categories:

• Exploration of the VR environment, during which the user rotates the head to observe the surroundings or to find the visual focus of the current stage.

• Rotation brought by visual cues, such as the tracking of moving objects.

• Rotation guided by sound cues, as indicated in a previous study[19].

(2) The scene-irrelevant head rotations were classified into the following two categories:

• Active movements with incentives from the user, such as the moving of the head for neck relaxation or movements performed when trying to reach a real-world object while wearing the HMD.

• Passive reactions toward an outside distraction, such as head jerks when patted on the shoulder.

Based on previous studies regarding eye-head movement patterns[15–17,20], we propose a correlation between the eye-head coordinate information and the scene-relevancy of a movement. To verify this assumption, we designed a pilot experiment that analyzes the various types of aforementioned actions for a better understanding of the identification task.

3.2 Preliminary experimentation

Both scene-relevant and scene-irrelevant head rotations are evoked by different types of stimuli. We adopted a movie-watching task for the best coverage of scene-relevant actions because most movies contain both visual and sound cues, as well as environment-exploring aspects. For scene-irrelevant head rotations, we adopted two types of possible actions in our experiments, which included adjusting the head for relaxation and turning the head when being patted on the shoulder, respectively, corresponding to active and passive irrelevant movements.

Authors invited 20 participants in our preliminary experiment, where they were asked to watch an immersive video[1]. During the experiment, they were asked to adjust their head when the instruction text appeared on the display, and turn their heads to respond to shoulder pats from the experimenter. We recorded the temporal sequences of their gaze shift and the degree of head rotation using the HTC VIVE Pro. After the experiment, we reviewed the recordings and manually marked the scene-irrelevant movements.

As illustrated in Figure 1, the eye and head rotations in scene-irrelevant movements occur at nearly the same time with quite similar speeds. For instance, regarding the positional differences between the eye and head in the eight consecutive frames after the movement occurred, 64 cases had an average of less than 5° and 112 cases had an average of less than 10° out of the 150 movements marked as scene-irrelevant. These characteristics agree with the existing psychological research results regarding non-predictive motions[15,16].

For the scene-relevant movements shown in Figure 2, we found that eye rotations often occur before head rotations. However, the exact latency between the eye and head rotations in scene-relevant situations appears to be inconsistent between the cases. Additionally, we noticed a number of cases in which the aforementioned eye-head coordination phenomenon does not appear to apply to rotations occurring at nearly the same time; in these cases, other characteristics were often present, such as earlier stopping times for the eye rotation. In general, patterns for scene-relevant movements vary and are difficult to detect using fixed algorithms.

Apart from the typical relevant and irrelevant samples, there were samples for which the patterns were difficult to recognize. There are approximately 35 cases in the 150 marked irrelevant movements that can be vaguely defined as difficult; for relevant movements, the number was difficult to determine because we could not exactly define the useful characteristics of scene-relevancy. Thus, we adopt machine-learning methods owing to their excellent ability to capture complex data dependencies and robustness.

3.3 Algorithm design

We modeled the problem described above as a binary classification task. Given an input that includes a

[1] https://www.youtube.com/watch?v=SZ0fKW5PttM

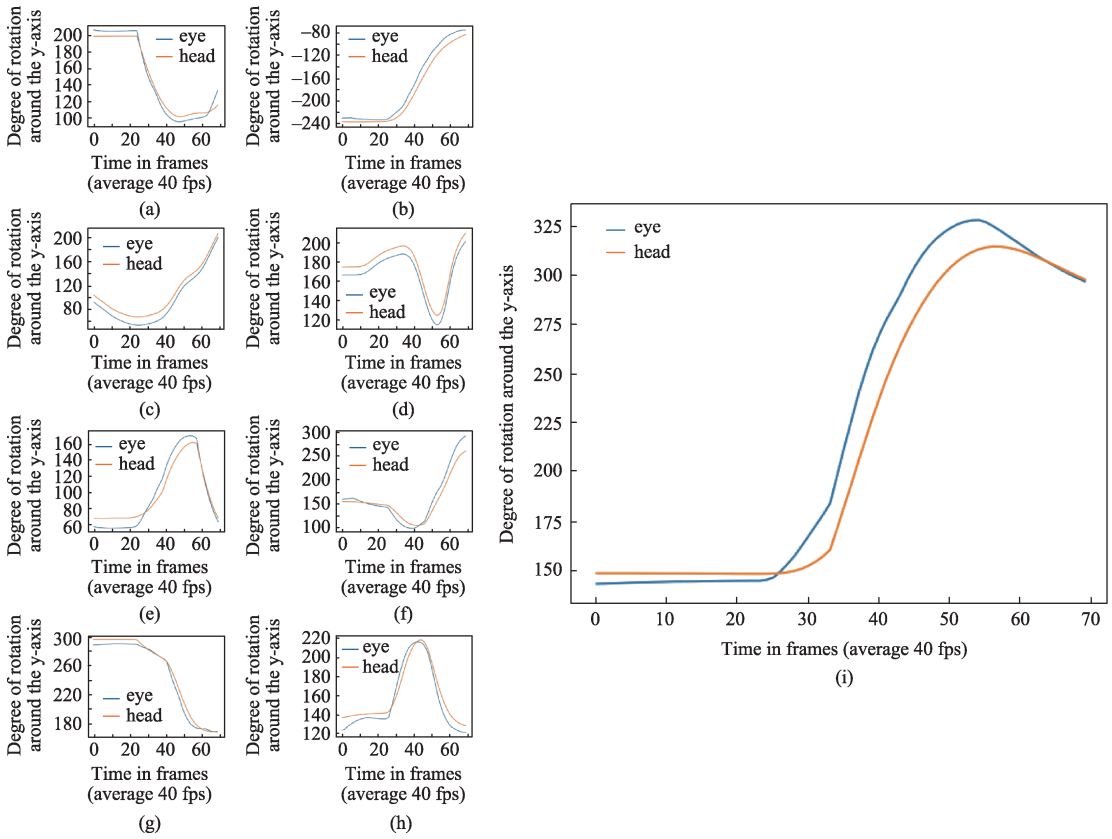

Figure 1 Nine examples of typical cases of scene-irrelevant movements, with the eye and head rotations always occurring at nearly the same time with similar speeds.

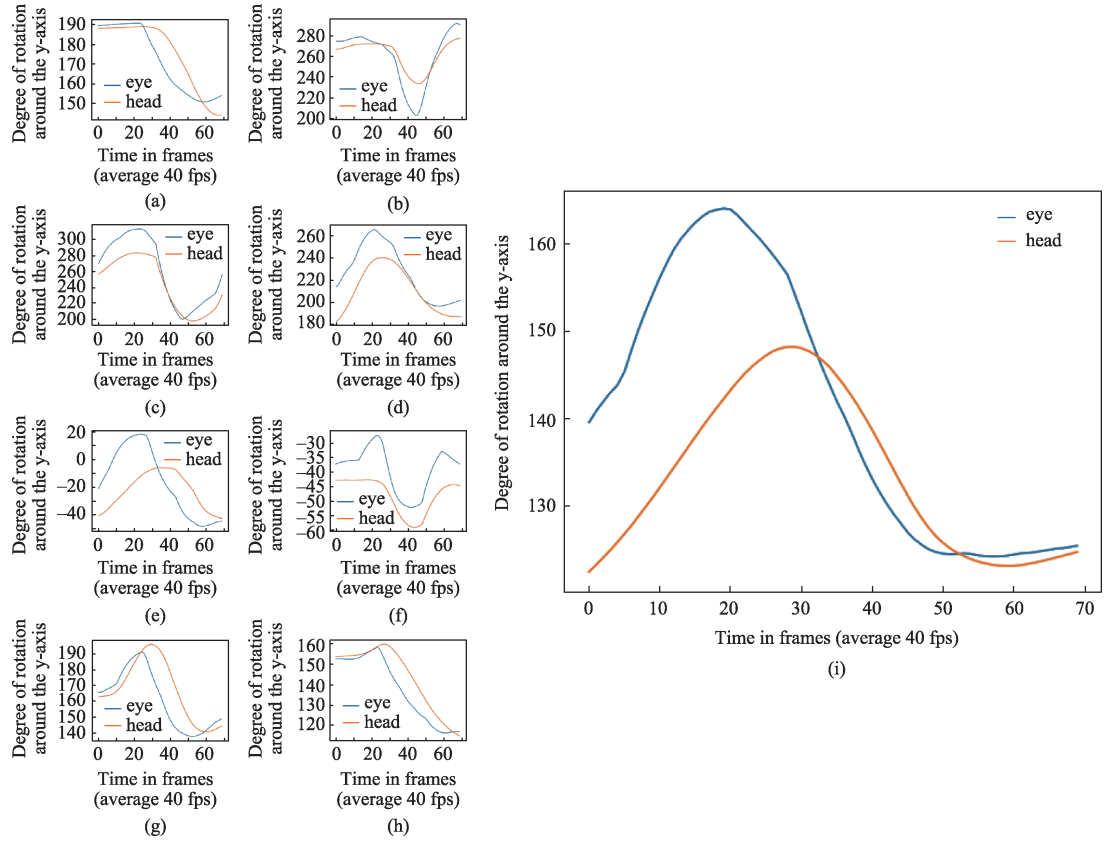

Figure 2 Nine examples of typical cases of scene-relevant movements, where eye rotations often occur before head rotations.

temporal eye and head orientation sequence denoted by $g_i, h_i \in \mathbb{R}^3$ at the i-th frame, and the timestamp $t_i \in \mathbb{R}$ of each frame, the output of the algorithm should be a temporal Boolean sequence indicating whether a scene-irrelevant movement occurs at each frame.

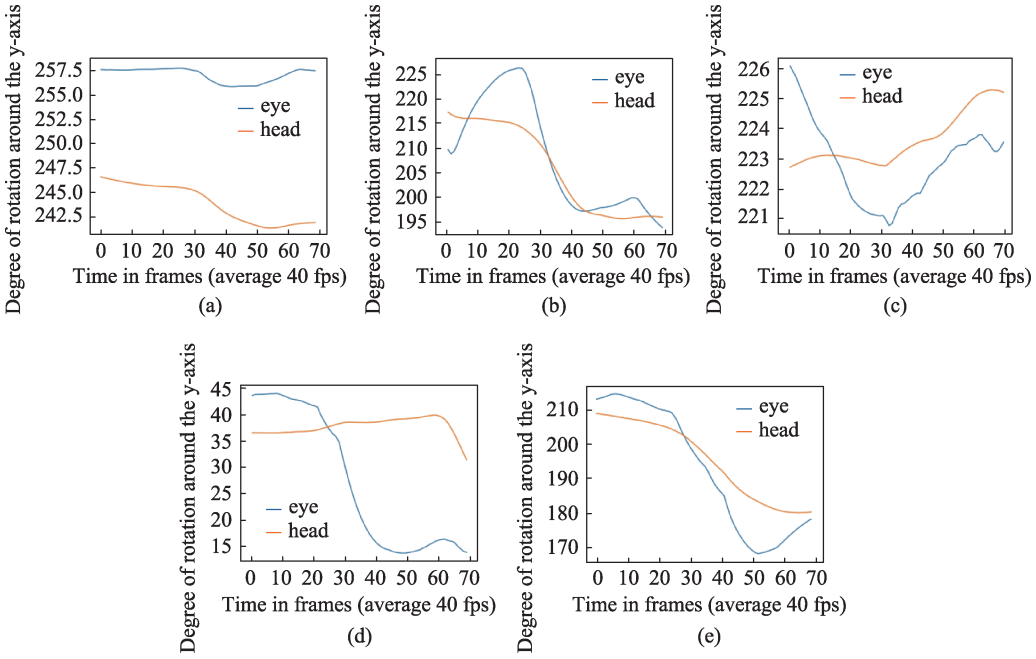

Figure 3 Five examples of difficult cases.

3.3.1 Input preprocessing

According to previous analysis, the algorithm should be ideally based on the movement information of the neck and head; therefore, we calculated the angular velocity of the eye and head at each frame. Specifically, we approximated the angular velocities E_i and H_i of the gaze and head orientations at the i-th frame using the following:

$$E_i = \frac{g_{i+1} - g_i}{t_{i+1} - t_i} \tag{1}$$

$$H_i = \frac{h_{i+1} - h_i}{t_{i+1} - t_i} \tag{2}$$

where g_i, h_i, and t_i are the eye, head, and timestamp data at the i-th frame, respectively, as previously indicated.

We classify the angular velocities of the eye and head in a sliding window on the input sequence. The sliding window is k frames wide, ranging from frames $(i - k, i]$, where k is a hyperparameter. The classifier should output a single Boolean for every time slice, with positive values denoting a scene-irrelevant movement. We concatenated the eye and head angular velocities of each frame inside the window as the input feature at frame I as follows:

$$In_i = \left[H_{i-k+1}^T, \cdots, H_i^T, E_{i-k+1}^T, \cdots, E_i^T\right]^T \in \mathbb{R}^{6k} \tag{3}$$

Although more information from the input data can be introduced into the later stages, such as the difference between the angles of gaze and the head orientation in the same frame, we found that they only have negative effects on the classification results in our experiments. The reason for this was believed to be the lack of a large dataset and the overfitting phenomenon.

3.3.2 Algorithm selection

While selecting from various machine-learning algorithms, we considered that first, our dataset is relatively small, and as such, the overfitting problem could occur. Second, the method must perform well under real-time circumstances, and therefore, it should be lightweight. In addition, regressors rather than classifiers should be adopted to better manage the different values of k and adapt to the training process, which will be explained later in this paper. The raw output of the regressor should be a value in the range [0, k], and converted into a binary classification output via a threshold δ during post-processing.

We tested a variety of methods, including the support vector regressor (SVR), multilayer perceptron regressor (MLPR), and k-nearest neighbor regressor (KNNR). SVR can be described as follows:

$$SVR(x) = \sum_{i=1}^{N} \alpha_i^* \exp\left(-\gamma \left|\left| x - x_i \right|\right|^2 \right) + b \tag{4}$$

where α_i^* and b are the learned parameters, N is the number of training samples, and γ is the reciprocal of the input dimension. The MLPR has two hidden layers with output sizes of 4 and 2, and ReLU activation layers after each one, trained with the limited-memory BFGS method and a maximum of 4000 iterations to ensure parameter convergence. It can be described as follows:

$$MLPR(x) = ReLU(W_2 \cdot ReLU(W_1 \cdot x + b) + b) \tag{5}$$

The KNNR can be described as follows:

$$KNNR(x) = argmax_{c_j} \sum_{x_i \in N_k(x)} I\left(y_i = c_j\right), i = 1,2,\cdots,N \tag{6}$$

where c_j are the categories denoting an integer in [0, k], N_k (x) is the set of the k-nearest neighbors determined by the Euclidean distance, and I() is the indicator function.

The hyperparameters of these methods, such as the learning rate in MLPR or the number of neighbors in KNNR, should be determined during the selection process.

4 Evaluation

In this section, we conduct a data collecting experiment to build our raw dataset, and then perform training and testing procedures on the dataset to determine the hyperparameters of the algorithm. The detection results were highly acceptable, which verified the previous analysis and design of the task.

4.1 Data collection

We first conducted an experiment to collect and annotate a sufficient amount of eye-head rotation data to learn the patterns of the eye-head movements during scene relevant/irrelevant actions. Participants were required to watch five 360-degree videos while sitting in non-rotatable chairs. While watching these videos, they were interrupted at random times and were required to respond to the interruptions. Certain interruptions were automatically executed by the system through text instructions, such as "Adjust head position" or "Turn Left", while others, such as shoulder-pats, were conducted by the experimenter. The ratios of the two stimulations were near 1∶1. The interruption, eye-head rotation data (binocular gaze direction and head orientation data) in the Euler angle, and time stamps were simultaneously recorded at a rate of approximately 80fps. At the same time, our system recorded a video of a participant's view on the VR display. We then annotated the starting point of the scene-irrelevant head rotations, which can be easily identified in the recorded video. Fifteen participants were involved in this experiment, and a total of 350min temporal sequences were collected. In a simple preprocessing step, we applied a Kalman filter to the raw orientation data to prevent hardware jitters from interfering with the following process.

Unfortunately, there are usually only a few scene-irrelevant movements compared to the number of frames in an experiment. This leads to a significant imbalance between the positive and negative samples. To resolve this issue, we first filtered out the negative samples in which the value of the angular velocity of the head orientation remained zero because we only triggered the classifier when detecting an apparent head movement. Then, we randomly down sampled the filtered negative samples to balance the positive and negative samples.

To generate the ground truth for the training process, it is necessary to provide each frame a positive or negative label regarding scene-irrelevancy, as the annotators only mark the start of each scene-irrelevant head movement in the raw dataset. For simplicity, we labeled the $[i,\ i+c)$ frames after an annotated frame i as positive, where c is a hyperparameter; all the other frames were labeled as negative. Then, we labeled the time slice as $(i-k,\ i]$ at i to be positive if the number of positive-labeled frames inside the window was more than a percentage threshold δ, as shown in Figure 4. Finally, we randomly split our dataset into training and testing sets in the ratio of 4∶1.

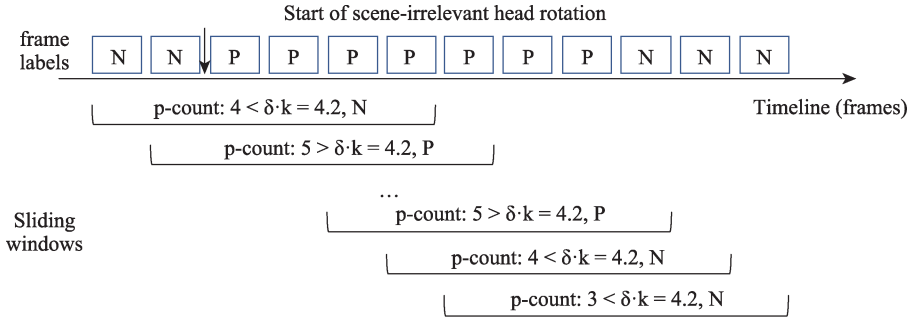

Figure 4 Time slice labeling scheme, with c=7, k=6, and δ=0.7.

4.2 Classification

We first experimented with different values of c and k on the KNNR because they affect the dataset at a relatively early stage in our implementation. For example, with each value of k, the length and number of time slices available changes, requiring the input-output pairs to be reorganized. The KNNR in this stage uses *n neighbors*=6, with δ=0.5. We found the classifier to be insensitive to c and k as the accuracy changes only 1% to 3% when c increases from 6 to 14 with k within the range $[c-1,\ c+1]$; as such, we set c=7 and k=6. Thus, the processed dataset comprised 4900 time slice samples, with the input to the regressor being a one-dim array of 36 features and the output being a single Boolean value indicating the action to be scene-irrelevant.

Next, a comparison between the methods was conducted to determine the suitable regressor, as well as the value of δ. We rough-tuned the MLPR and KNNR hyperparameters at δ=0.5. For the MLPR, the weight decay for the limited-memory BFGS method was 0.0001; for the KNNR, *n neighbors*=6. As this was a fairly standard case of the binary classification problem, we mainly considered the final accuracy and f1-score during the selection process, with regard to the precision and recall scores. As illustrated in Table 1, all the methods performed optimally when $\delta \in [0.4,\ 0.6]$. Finally, the KNNR with δ=0.4, an accuracy of 0.8205, and an f1-score of 0.7602 was chosen for the view-following technique. Notably, combining the three methods into a voting regressor did not lead to an increase in performance.

Regarding the KNNR, we tuned the hyperparameter *n_neighbors*. Finally, we chose *n_neighbors*=4, with an accuracy of 0.8487 and f1-score of 0.8090.

The success of the classification task further ensures that a relationship between the eye and head

Table 1 Comparison of the methods

Method	δ	Accuracy	Precision	Recall	f1
Support vector regressor	0.4	0.8064	0.6738	0.8295	0.7436
	0.5	0.7576	0.6764	0.1144	0.1957
	0.6	0.7435	0.6666	0.0099	0.0196
Multilayer-perceptron regressor	0.4	0.7884	0.6542	0.7954	0.7179
	0.5	0.773	0.5419	0.7711	0.6365
	0.6	0.7743	0.5576	0.6019	0.5789
K-nearest neighbor regressor	0.4	0.8205	0.6937	0.8409	0.7602
	0.5	0.8192	0.6209	0.7661	0.6859
	0.6	0.8397	0.7065	0.6467	0.6753

movements exists in the coordinated gaze shift in the VR environments, which can be utilized to distinguish scene-irrelevant head movements from relevant ones. To verify the applicability of the identification results, we designed a view-following technique.

5 Application and user study

In this section, we present the design and implementation of a practical view-following technique to utilize the detection results. We conducted a controlled experiment to verify the feasibility and effectiveness of VR movie-viewing experiences. We found that the technique provides a better user experience by reducing the impact of user scene-irrelevant movements, resulting in a higher continuity without increasing the sickness or breaking immersion.

5.1 Implementation of view-following technique

We propose a practical technique for a movie-watching situation. Our design ignores detected scene-irrelevant movements because they are not intended for the VR environment. This allows users to keep track of their original targets while performing those actions, and also grants them the ability to smoothly switch from one position to another, for instance from a sitting position to a lying position.

In our implementation, the video was projected onto a skybox material, which can be considered as the inner surface of a sphere whose center is the viewer. The video usually remained fixed to the VR environment (world-locked); however, it can also be set to rotate along with the head (head-locked) when needed, thus achieving the "view following" effect. This technique centers around the scene-irrelevant detection algorithm. Real-time eye and head orientation and timestamps were collected from the HMD. When the video is in a world-locked state, head rotation information from the HMD was first used to determine whether the head was moving. When an apparent movement occurred, the relevant stored information was fed as a sequence into the algorithm. If the algorithm detected a scene-irrelevant movement, the video was set to head-locked. It remains in the head-locked state until there no apparent movement from the head was noted; then, it returned to the world-locked state. Note, switching between states may likely repeatedly occur in practice; for instance, when a person jerks his head toward a distraction and quickly jerks back.

5.2 Experiment design

To evaluate the effectiveness of the view-following technique, we selected three evaluative dimensions: *continuity*, *sickness*, and *immersion*. *Continuity* refers to the level at which users can keep track of the

subject observed while completing the assigned head movements. *Sickness* refers to the level of users showing symptoms of 3D motion sickness, including vertigo, nausea, vomiting, and headache in VR. *Immersion* refers to the level where the users are absorbed in the virtual world and storytelling without feeling distracted, disengaged, or lost. A good continuity, low sickness, and unbroken immersion contributes to a better VR experience.

The technique is supposed to provide a better user experience by reducing the impact of user scene-irrelevant movements. Correspondingly, we designed a 2×2 mixed-design study to evaluate this technique. We selected two 360-degree videos and used them to conduct two groups of experiments, A and B. In these experiments, the participants were required to watch the videos with one applying the technique and the other not, as can be observed from Table 2. Here, we did not specify the order in which the participants watched the two videos because we asked the users to take a 5-min break between the two experiments, eliminating the risk of sickness increasing as the VR watching time increases.

Table 2 Mixed experience set

Groups	Video1: Invasion	Video2: Rain or Shine
A B	with our technique without our technique	without our technique with our technique

According to the goal of our study, our main hypotheses for the experiment are as follows:

• H1: The level of continuity is higher in VR movies with our gaze-guided view control technique than without in circumstances containing scene-irrelevant head movements.

• H2: Whether the gaze-guided view control technique is applied to the VR movies does not invoke significant differences in the level of sickness.

• H3: Whether the gaze-guided view control technique is applied to the VR movies does not invoke significant differences in the level of immersion.

• H4: Different videos with the same treatment (with or without the technique) do not present differences in the levels of continuity, sickness, or immersion in the VR experience.

We classified the four experiments presented in Table 2 into two groups: video1 and video2 as between-subjects. Each video had the following two conditions: one in which our technique is used and the other in which our technique is not used, which were within-subjects. H1, H2, and H3 were hypotheses regarding the within-subjects, whereas H4 is a hypothesis regarding between-subjects. Significant tests were conducted with p-values calculated to verify the aforementioned hypotheses.

The two 360-degree movies used in our study were *Invasion* from Baobab Studios[2] and *Rain or Shine* from Google Spotlight Stories[3] (Figure 6), which are available online for free. They are short (6min and 5.5min, respectively), digitally animated narrative 360-degree videos with interesting and captivating plots, to try and keep the participants' focus on the subject in the video as much as possible; moreover, the movies require the participates to make a variety of eye and head movements, thereby making the videos perfectly suitable for the testing of the detection algorithm.

Figure 5 Experimental environment. Left: a participant watching a movie clip. Top right: a line chart for experimenters to keep track of eye and head rotations of the participant. Bottom right: a scene from the movie.

[2] https:/Twohttp://www.baobabstudios.com

[3] https://atap. google. com/spotlight stories/

5.3 Apparatus and participants

The proposed technique was implemented using the Unity3D game engine version 2019.4.6. The HTC VIVE Pro Eye (embedded with an eye tracker) was used as an HMD for head and eye tracking. The application ran on a desktop computer with an Intel Core I7 processor and Nvidia GeForce GTX 1080Ti graphics card at a speed of approximately 80 frames per second. With the eye-tracking module running at 120Hz and the head tracking module running at 90Hz, we were able to record the head and eye rotation data per frame through Unity3D. Participants were required to sit in a non-rotatable chair during the process, which was consistent with the data collection experiment we conducted previously.

Figure 6 Example of still frames from the videos used (single frame used under fair use for research and demonstration purposes)

Figure 6 Example of still frames from the videos used (single frame used under fair use for research and demonstration purposes).

Twenty participants (5 females and 15 males) were involved in the user evaluation study, whose ages were mainly between 21−25. All the participants reported that they had previous experience with VR, including playing 3D or VR games and watching 360-degree videos. Regarding the assessment of 3D motion sickness symptoms, such as vertigo, nausea, vomiting, and headache, 9 participants showed no symptoms of motion sickness, 10 showed mild symptoms of motion sickness or occasional carsickness, and one developed symptoms of motion sickness after 10min of playing 3D games, according to their self-reported information.

5.4 Procedure

At the beginning of the experiment, each participant was asked to complete a background questionnaire to provide information regarding their age, gender, experience with VR, and symptoms of 3D motion sickness. The main task of each participant in the experiment was to watch two VR videos, where the view-following technique was applied in one, and not, in the other, and then to provide feedback afterwards regarding both the cases.

Before watching, the participants were informed that they would be interrupted several times while they would be watching the videos, with external distractions, such as text instructions from the HMD, experimenter's actions, or questions. They were also asked to respond to the interruptions by completing the assigned actions, such as by turning their head, adjusting their head for comfort, and answering the experimenter's questions. Meanwhile, they were required to keep track of the subject which they had been concentrating on. Each participant underwent a calibration process of an eye-tracking software before watching the videos to ensure that the movie-watching quality and the classification accuracy of our technique were desirable.

Each participant watched the two VR movies; our technique was applied to one of these videos according to the group to which the participant was assigned, as described in Section 5.2.

After watching the movies, the participants were asked to rate the continuity, sickness, and overall

immersion with regard to their viewing experience of the two videos on a 5-point Likert scale. The experimenter then conducted an informal interview to collect other details that the participants felt noteworthy in terms of their experience.

5.5 Experiment results

After collecting the data from the questionnaires, we analyzed the frequency counts of each evaluative scale for each video with or without our technique being used. We prepared stacked bar charts to visualize the percentage distributions (Figures 7, 8, and 9). We then conducted significant tests with scores from the comparison groups specified in Section 4.2 to verify our hypotheses.

To verify hypothesis H1, we analyzed the scores of continuity for the watching experiences with and without our technique using the Wilcoxon signed rank test, finding a significant difference (p=0.0068<0.05). There is a rightward shift between the rating distribution for the watching experiences with our technique (Mean=4, SD=0.21) and without our technique (Mean=3.2, SD=1.01), as shown in Figure 7, which indicates a significant improvement in the continuity in VR movie-watching experiences when unintended movements occur. To verify hypothesis H2, we analyzed the sickness data when watching videos with our technique and videos without our technique using the Wilcoxon signed rank test without finding a significant difference (p=0.8738>0.05). The results indicate that there is no significant difference in the level of sickness between the videos with and without our technique, which validates that our gaze-guided view control technique will not increase sickness in VR experience.

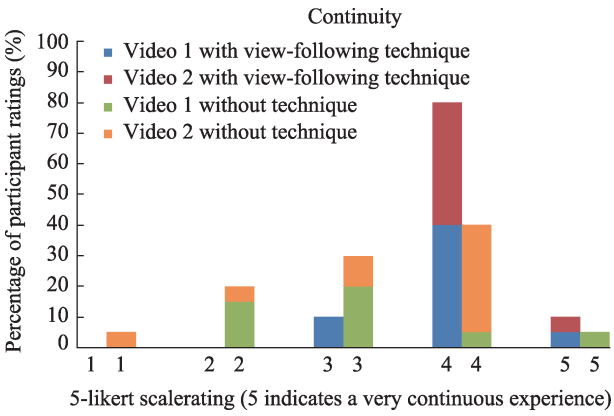

Figure 7 Percentage distribution of ratings on continuity.

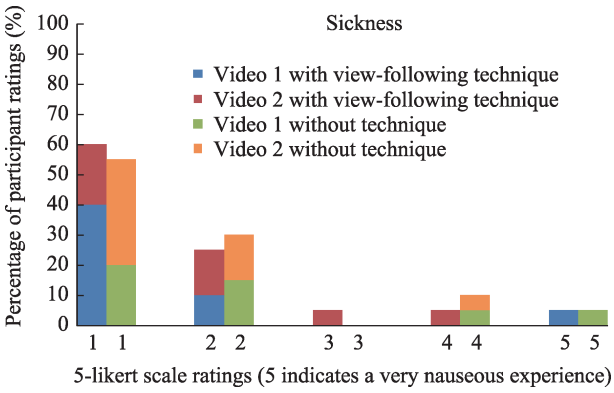

Figure 8 Percentage distribution of ratings on sickness.

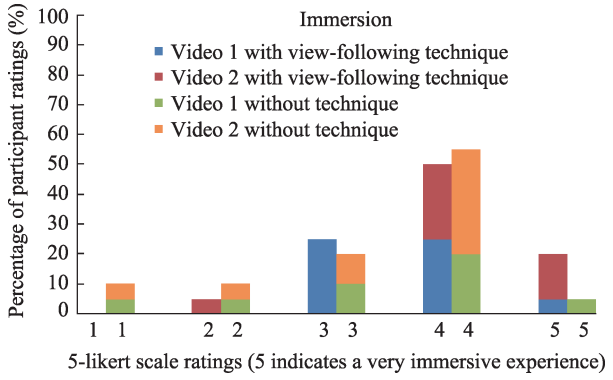

Figure 9 Percentage distribution of ratings on immersion.

To verify H3, we analyzed the immersion data from watching videos with our technique and videos without our technique using the Wilcoxon signed rank test without finding a significant difference (p=0.0840>0.05). Note, the out technique (Mean=3.85, SD=0.66) provides the ratings of a rightwards shift than without (Mean=3.35, SD=1.18), almost managing to significantly increase immersion. This validates that our technique does not break immersion, and slightly increases it.

To verify H4, we compared the data of the three criteria for the two movies with the same treatment using the Wilcoxon signed-rank test. For the scores of watching different movies without using our technique, we did not find significant differences in continuity (p=0.3438>0.05), sickness (p=0.2656>

0.05), or immersion (p=0.7500>0.05). For the continuity data of watching the two movies using our technique, no significant differences in continuity (p=1.00>0.05), sickness (p=0.3438>0.05), or immersion (p=0.3438>0.05) were found once again. This indicates that the different videos will not cause different feelings during unintended head movements in the VR experience, as we supposed.

During the informal interview after the experiment, the participants mostly gave positive feedbacks on the technique. Seventeen out of 20 participants noticed the gaze-guided view control technique we applied to one of the movies, of which 16 reported that this increased the overall immersion. A few users reported that they spent less time re-seeking the character they had been watching before the interruption, which reduced the discomfort of losing focus in scene-irrelevant head movements. However, a few participants claimed to be confused about the scene orientation after the view following occurred; thus, perhaps a more user-friendly action can be presented once a scene-irrelevant movement is detected. Nonetheless, the results of the user study confirmed the effectiveness of our technique.

6 Conclusion and limitations

To the best of our knowledge, this is the first study that explored the detection of scene-irrelevant movements and the application of eye-head coordinated movement patterns in view control. Our study reveals that scene-irrelevant head rotations can be effectively detected through eye-head coordination behaviors. The study results suggest that eye-head coordination behaviors can be utilized for other tasks, such as user intention detection. We also proposed a view-following technique to utilize the detection of scene-irrelevant head rotation over a common movie-watching VR scenario. The study results also indicate that our technique improves the continuity of the narrative experience without increasing the sickness level or decreasing the immersion level, demonstrating that it is advisable to manage irrelevant movements for better user experience.

However, our study has a few limitations. Currently, we model the problem of irrelevancy detection as a binary classification problem. However, we acknowledge that existing eye-head coordination studies[1,10,16,20] have indicated the existence of several patterns of eye-head movements in various conditions and tasks. It is possible that the problem can be modeled better in the form of a multi-class classification task. Moreover, larger datasets with more varied VR activities are required to promote the development of eye-head coordination studies. The detection of scene-irrelevant actions and the utilization of eye-head coordinated movements have a significant potential to be studied and applied in VR scenarios. We look forward to further studies in this direction.

Declaration of competing interest

We declare that we have no conflict of interest.

References

1 Kavanagh S, Luxton-Reilly A, Wuensche B, Plimmer B. A systematic review of virtual reality in education. Themes in Science and Technology Education, 2017, 10(2): 85−119

2 Mahrer N E, Gold J I. The use of virtual reality for pain control: a review. Current Pain and Headache Reports, 2009, 13 (2): 100−109
DOI:10.1007/s11916-009-0019-8

3 LaValle S M, Yershova A, Katsev M, Antonov M. Head tracking for the oculus rift. In: 2014 IEEE International Conference on Robotics and Automation (ICRA). Hong Kong, China, IEEE, 2014, 187−194
DOI:10.1109/icra.2014.6906608

4 Gandrud J, Interrante V. Predicting destination using head orientation and gaze direction during locomotion in VR. In:

Proceedings of the ACM Symposium on Applied Perception. Anaheim California, New York, NY, USA, ACM, 2016, 31–38
DOI:10.1145/2931002.2931010

5 Wille M, Grauel B, Adolph L. Strain caused by head mounted displays. Proceedings of the Human Factors and Ergonomics Society Europe, 2013, 267–277

6 Waterworth E L, Waterworth J A. Focus, locus, and sensus: the three dimensions of virtual experience. Cyberpsychology & Behavior, 2001, 4(2): 203–213
DOI:10.1089/109493101300117893

7 Garau M, Friedman D, Widenfeld H R, Antley A, Brogni A, Slater M. Temporal and spatial variations in presence: qualitative analysis of interviews from an experiment on breaks in presence. Presence, 2008, 17(3): 293–309
DOI:10.1162/pres.17.3.293

8 Oh C, Herrera F, Bailenson J. The effects of immersion and real-world distractions on virtual social interactions. Cyberpsychology, Behavior, and Social Networking, 2019, 22(6): 365–372
DOI:10.1089/cyber.2018.0404

9 Doshi A, Trivedi M M. Head and eye gaze dynamics during visual attention shifts in complex environments. Journal of Vision, 2012, 12(2): 9
DOI:10.1167/12.2.9

10 Pfeil K, Taranta E M II, Kulshreshth A, Wisniewski P, LaViola J J. A comparison of eye-head coordination between virtual and physical realities. In: Proceedings of the 15th ACM Symposium on Applied Perception. Vancouver British Columbia Canada, New York, NY, USA, ACM, 2018, 1–7
DOI:10.1145/3225153.3225157

11 Zank M, Kunz A. Eye tracking for locomotion prediction in redirected walking. In: 2016 IEEE Symposium on 3D User Interfaces (3DUI). Greenville, SC, USA, IEEE, 2016, 49–58
DOI:10.1109/3dui.2016.7460030

12 Freedman E G. Coordination of the eyes and head during visual orienting. Experimental Brain Research, 2008, 190(4): 369–387
DOI:10.1007/s00221-008-1504-8

13 Dodge R. The latent time of compensatory eye-movements. Journal of Experimental Psychology, 1921, 4(4): 247–269
DOI:10.1037/h0075676

14 Clément G, Reschke M F. Compensatory eye movements. Springer Science & Business Media, New York, NY, 2010, 163–188

15 Zangemeister W H, Stark L. Gaze latency: Variable interactions of head and eye latency. Experimental Neurology, 1982, 75(2): 389–406
DOI:10.1016/0014-4886(82)90169-8

16 Bizzi E. The coordination of eye-head movements. Scientific American, 1974, 231(4): 100–106

17 Mourant R R, Grimson C G. Predictive head-movements during automobile mirror-sampling. Perceptual and Motor Skills, 1977, 44(1): 283–286
DOI:10.2466/pms.1977.44.1.283

18 Di Girolamo S, Picciotti P, Sergi B, Di Nardo W, Paludetti G, Ottaviani F. Vestibulo-ocular reflex modification after virtual environment exposure. Acta Oto-Laryngologica, 2001, 121(2): 211–215
DOI:10.1080/000164801300043541

19 Rothe S, Hußmann H, Allary M. Diegetic cues for guiding the viewer in cinematic virtual reality. In: Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology. Gothenburg Sweden. New York, NY, USA, ACM, 2017
DOI:10.1145/3139131.3143421