Lost in Style: Gaze-driven Adaptive Aid for VR Navigation

Rawan Alghofaili
George Mason University
Fairfax, Virginia

Yasuhito Sawahata
Japan Broadcasting Corporation
Setagaya, Tokyo, Japan

Haikun Huang
University of Massachusetts-Boston
Boston, Massachusetts

Hsueh-Cheng Wang
National Chiao Tung University
Hsinchu, Taiwan

Takaaki Shiratori
Facebook Reality Labs
Pittsburgh, Pennsylvania

Lap-Fai Yu
George Mason University
Fairfax, Virginia

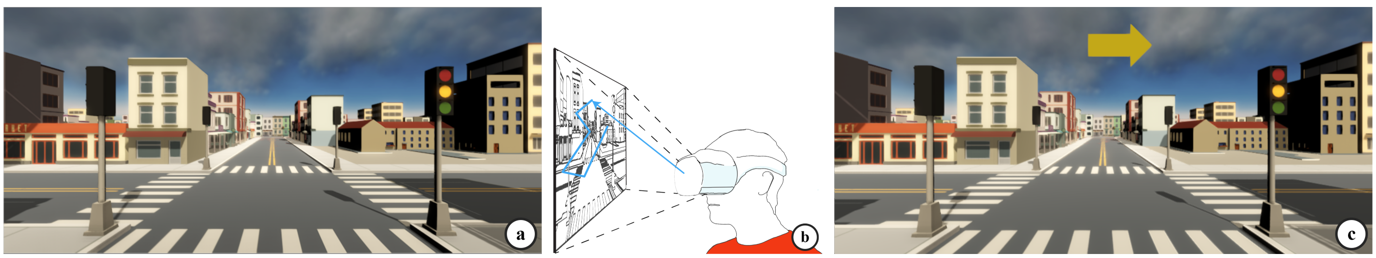


Figure 1: (a) While navigating a scene in virtual reality, (b) the user's gaze sequence can indicate his/her need for navigation help and (c) an aid is displayed accordingly.

ABSTRACT

A key challenge for virtual reality level designers is striking a balance between maintaining the immersiveness of VR and providing users with on-screen aids after designing a virtual experience. These aids are often necessary for wayfinding in virtual environments with complex paths.

We introduce a novel adaptive aid that maintains the effectiveness of traditional aids, while equipping designers and users with the controls of how often help is displayed. Our adaptive aid uses gaze patterns in predicting user's need for navigation aid in VR and displays mini-maps or arrows accordingly. Using a dataset of gaze angle sequences of users navigating a VR environment and markers of when users requested aid, we trained an LSTM to classify user's gaze sequences as needing navigation help and display an aid. We validated the efficacy of the adaptive aid for wayfinding compared to other commonly-used wayfinding aids.

CHI 2019, May 4–9, 2019, Glasgow, Scotland Uk

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300578

CCS CONCEPTS

• **Human-centered computing → Virtual reality**;

KEYWORDS

Games/Play, Virtual/Augmented Reality, Eye Tracking

ACM Reference Format:
Rawan Alghofaili, Yasuhito Sawahata, Haikun Huang, Hsueh-Cheng Wang, Takaaki Shiratori, and Lap-Fai Yu. 2019. Lost in Style: Gaze-driven Adaptive Aid for VR Navigation. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland Uk.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3290605.3300578

1 INTRODUCTION

With the introduction of relatively inexpensive virtual reality (VR) head mounted displays that have eye-tracking capabilities like the FOVE, the barrier of entry to utilizing eye-tracking data for virtual reality applications has been lowered significantly. Conducting eye-tracking experiments in virtual reality becomes significantly simpler, fast tracking behavioral and cognitive studies.

Additionally, the availability of these devices and their corresponding data sets provide researchers and developers with ample opportunities for creating more immersive experiences for VR game players by utilizing eye-tracking technologies.

In virtual experiences, users often face difficulties in exploring and navigating a virtual environment. This often is frustrating and weakens the immersive factor that makes virtual experiences so enticing. These feelings of 'needing navigation aid' often go unnoticed by developers until the user issues a complaint about the navigation difficulty. This leads developers to make difficult design choices to incorporate permanent navigational guides to their interfaces, which may come in forms like mini-maps, arrows or overlaid paths. All of these navigational guides may take up considerable screen space and could be visually unappealing or distracting to embed in the VR interface.

We propose a novel approach to predict VR users' need for navigation aid by analyzing their gaze patterns. Figure 1 shows an illustration. To experiment with our approach, we collected gaze and head motion data from subjects recruited to complete a challenging search task in a virtual environment. This dataset was used to train a gaze sequence classifier to predict the player's need for navigation aid when navigating a virtual environment. We discuss the model's prediction accuracy and present several applications of the model as adaptive navigation aids. We explore our adaptive navigation aids' abilities in effectively guiding the player to navigate a virtual environment, as well as using gaze heatmaps to compare the player's visual attention to the surroundings when using our adaptive aid versus that when using a conventional permanent on-screen aid.

The major contributions of our work include:

- Proposing a novel data-driven approach to train a sequence classifier to predict navigation aid need of users in VR using their gaze and head movements.
- Based on the sequence classifier, adaptively displaying a navigation aid to users when their gaze sequences indicate that they are feeling in need of aid.
- Validating the effectiveness of our adaptive aids compared to permanent aids via a number of human user experiments conducted in virtual environments.

2 RELATED WORK

Eye-Tracking

Eye-tracking and gaze patterns have been employed for classifying human behavior and cognition in addition to improving user experience. For example, Sanches et al. [27] analyzed the gaze behavior of subjects to estimate their understanding of texts while reading. Lustig et al. [17] utilized an LSTM to classify subjects' reading behavior as dyslexic. Similarly, we employed an LSTM to classify gaze sequences of a user feeling in need of guidance while navigating a virtual environment. Gaze patterns have also been analyzed for improving user experience. For example, Turner et al. [30] demonstrated how gaze could be used for manipulating objects on a display.

The relationship between gaze fixation flow and spatial navigation was studied in [29, 32] with a focus on Alzheimer's patients. Burch [1] analyzed visual attraction on metro maps using gaze fixations. Piccardi et al. [24] found a relationship between the navigational styles of subjects and their gaze patterns while learning environment maps. We analyzed gaze patterns for classifying navigation guidance need in virtual environments.

Virtual Reality

Interaction: VR and AR devices have become much more accessible thanks to the advancement and popularity of consumer-grade VR and AR devices such as the Oculus Rift, HTC Vive and Microsoft Hololens. Studying interaction within those mediums has become a focal point and an open research problem for the HCI community. Some of the many interactive applications in VR include VR story board design tools [10], VR/AR conferencing systems [21] and novel tools for reviewing and editing VR videos [20].

More specifically, VR game interactions have been a key target of HCI researchers. Cheng et al. [3] presented a mutual actuation solution by using one VR players' actions in a VR game as haptic feedback for another player. Lopes et al. [16] introduced a haptic feedback solution to VR games by simulating physical impact while playing a video game using an electronic bracelet the authors designed. Like these works, our approach aims to improve the interactivity of VR by the use of an adaptive navigation aid.

Navigation: Our approach aims at enhancing navigation in virtual environments. Previous works [7, 8, 19, 25] explored adaptively expanding a virtual scene to facilitate VR locomotion in small physical spaces. Gandrud and Interrante [6] utilized a user's gaze and head motion to infer his navigation direction in VR, while Lee et al. [14] compared the visual attention of players navigating in a first person perspective with players playing in a third person perspective as avatars.

Saha et al. [26] conducted a VR experiment to study subjects' negative emotions when presented with the wrong service in a virtual supermarket. They guided the subjects to their destination with a navigation assistant, which was realized in the form of an embedded path similar to the path we used in our data collection experiments (see Figure 3). Moller et al. [18] utilized an on-screen arrow navigation aid much like ours in their AR and VR indoor navigation application. Darken et al. [4] studied the performance of a variety of VR navigation tools including: arrows, mini-maps, landmarks and embedded paths. They concluded that arrows are ineffective on their own as navigation aids and should be complemented with additional hints like landmarks. We added landmarks in our virtual space to ensure that the conclusion of our user study is not skewed by the inability of arrows to guide users.

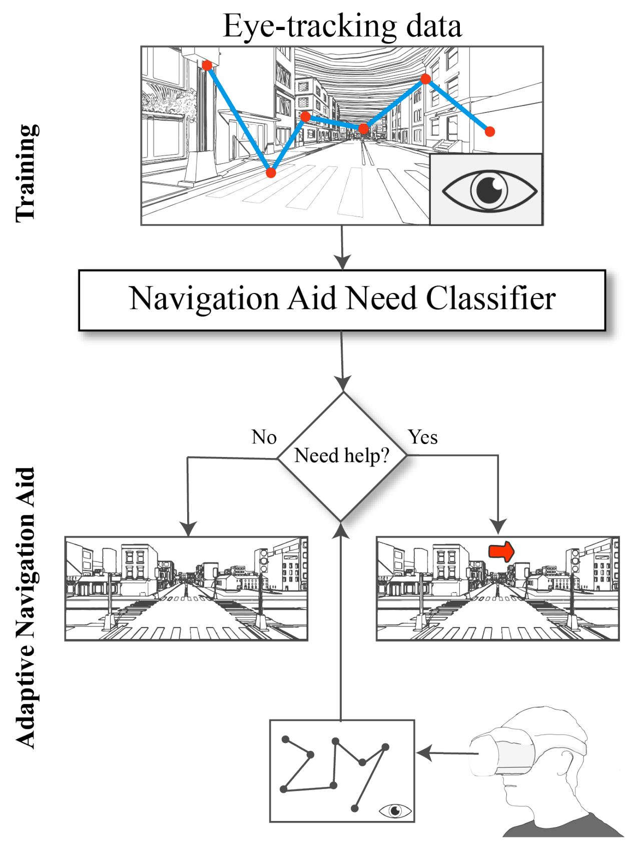


Figure 2: Overview of our approach.

Eye-tracking in HMD: Several attempts have been made to enable eye-tracking in head-mounted displays. Pai et al. [22] experimented with using gaze depth for interaction. They used an Oculus Rift HMD with an added eye-tracker designed by Pupil Labs. Shimizu et al. [28] modified the Google Cardboard by using an Electrooculography (EOG) module extracted from an AR device to pair the Cardboard with eye-tracking capabilities. We used an eye-tracking VR headset in conducting our experiments. Our HMD (FOVE), however, was pre-designed with an IR eye-tracker. Please refer to Velloso et al. [31] for several gaze-based game mechanics based on the FOVE.

3 OVERVIEW

Figure 2 shows an overview of our approach. By using eye-tracking data collected from subjects navigating a virtual environment, we trained a classifier to predict user's need for a navigation aid as a result of navigating a convoluted virtual environment. Essentially, our classifier learned gaze patterns indicating users will most likely request help in finding their way. Using this classifier, we experimented with adaptively displaying navigation aids to users. Finally we demonstrated the effectiveness of our aid by conducting a user study comparing traditional navigation aids to our adaptive approach.

Figure 3: Participants of our data collection experiment completed missions while wearing the (b) FOVE eye-tracking virtual reality headset. Their position and gaze data were recorded and later used to train our LSTM. (a) The red dot—not shown to the participant during gameplay— depicts this participant's gaze point, and the yellow navigation aid shows the closest path to the destination.

4 TRAINING DATA COLLECTION

In order to train a time-series classifier to classify gaze patterns of users as "needing" or "not needing " navigation help, we collected gaze data from participants navigating a VR space. We preprocessed this data set to remove noise and used a windowing approach to produce labeled gaze sequences to feed the classifier.

Participants: We recruited 22 college and graduate students with ages ranging from 18 − 30 to participate in our IRB-approved experiments.

Setup: To navigate in the environment, each participant wore a FOVE virtual reality headset which tracks his/her head orientation and gaze. The FOVE had a built-in infrared eye-tracking system that operates with a frame rate of 120 fps. An Internal Measurement Unit (IMU) was used to track the head orientation and an infrared sensor was used to track the user's gaze. It displayed visuals at a frame rate of 70 fps, while our program sampled data at 50 fps. Users controlled their locomotion using a game controller. Figure 3 shows the FOVE headset and an example of the display shown to participants.

Tasks: To collect the data set for training our classifier, we designed a 3D virtual city (*VR City*) in Unity complete with mock supermarkets, auto-repair shops, gas stations, etc. The layout of *VR City* is shown in Figure 4. Each participant completed a total of 4 tasks by navigating the city in virtual reality. The participant needed to find a fruit in each task.

In the first half of each task, we placed each participant at a randomly selected starting point from a pool of 5 pre-specified locations and asked him/her to find a fruit (apple, strawberry, banana or kiwi on each). Note that the closest path to the fruit *is shown* to the participant to help him/her find the fruit, as depicted in Figure 3.

After becoming familiar with the route by completing the first half (from the starting point to the fruit) which served as

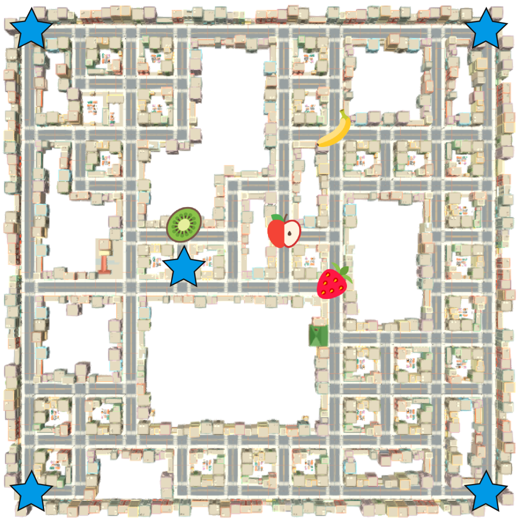


Figure 4: *VR city*'s layout. The participants were asked to navigate the virtual city in the training data collection sessions. The stars show the stating points' locations, while the other icons show the fruits' locations.

a training task, in the second half of each task, the participant was asked to return to the same starting point selected in the first half. However, this time, the path *was not shown* to the participant unless it was triggered it by pressing a button on the controller. If the path was triggered to be shown, the participant needed to wait for at least 30 seconds before he/she could trigger the path again. This time constraint was designed to avoid participants abusing the hint button, and to ensure that the hint was most likely displayed only when the participant needed it, reducing the likelihood of adding false positives to the training set. To further push participants to feel confused, we placed a barrier in a random location on the path they followed in the first half of the session while looking for the fruit. Figure 5 provides an overview of our tasks.

VR City was designed to be maze-like and difficult to navigate because we intentionally wanted participants to feel confused, thus pressing the hint button intermittently and providing us with positive samples (corresponding to navigation aid needs) to train our classifier.

5 NAVIGATION AID NEED PREDICTION

Data Processing

In each task a participant completed, we collected a time series of gaze angles, along with markers of whether the navigation aid was shown to the participant or not. We computed the gaze angle as follows:

$$\theta_t = \mathrm{acos}(\hat{\mathrm{g}}_t \bullet \hat{\mathrm{h}}_t), \quad (1)$$

where the the gaze angle θ_t at time t is computed using the inverse cosine of the dot product of the normalized gaze

Figure 5: We asked participants to finish tasks during the data collection phase. (a) The participants were dropped at the starting point. (b) They were asked to follow the path to (c) a fruit. After finding the fruit, participants were asked to find their way back to the starting point. The path was hidden from the participants unless triggered by pressing a button. (d) Barriers were randomly placed along the path back to the starting point to intentionally confuse the participants.

direction vector $\hat{\mathrm{g}}_t$ and normalized head forward direction vector $\hat{\mathrm{h}}_t$ at time t. Figure 6 shows an illustration of our gaze angle computation. In our experiments, we found that the gaze angle was sufficient to predict navigation aid need using our model. To better generalize our approach, we avoided including any features that could be scene specific (e.g., looking at buildings, turns, ground).

The series of gaze angles computed for a participant's session comprise a gaze session instance. We standardized our gaze session instances by subtracting their mean and scaling them to unit variance. We padded gaze instances that are shorter than the longest gaze instance with zero padding. This padding was only done for normalizing our data set and was removed before advancing to any of our other preprocessing steps.

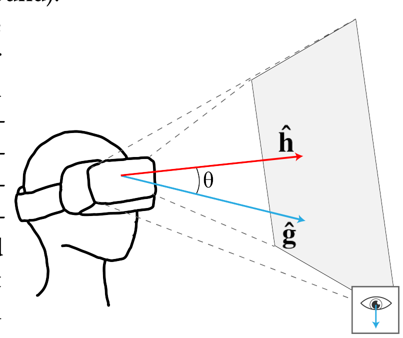


Figure 6: Computing the gaze angle.

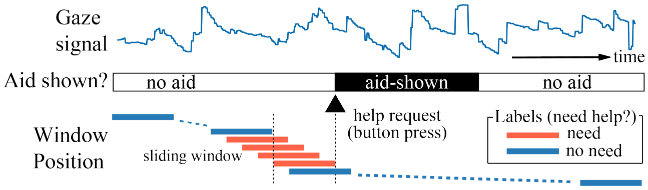


Figure 7: Windowing and labeling of gaze angle sequences. More details are provided in the Data Processing section.

Noise Reduction: Due to calibration errors for some participants, the FOVE eye-tracker produced some noisy sequences. Thus, it was necessary to smooth the sequences after standardization. We used a moving average filter with a window size of 50 time frames (1 second) to smooth out any noise in the gaze sequences.

Windowing: After preprocessing the gaze samples we used a sliding window to produce the final sequences for classification. We found that using a window of size 350 time frames provided us with the best trade-off between accuracy and the shortest possible window for prediction. Each 350 sized sequence gave us data from about 7 seconds of game-play. Figure 7 illustrates our windowing and labeling method.

Labeling: We labeled the windowed gaze sequences according to the navigation aid marker previously mentioned. Subjects likely felt a need for directional help prior to pressing the hint button, i.e., before the navigation aid marker indicated that the path was shown (see Figure 7). Consequentially, we labeled all the windowed gaze sequences within a window's width (or 350 time frames) before a navigation marker as "positive" (need for navigation aid). The remaining windowed gaze sequences were labeled as "negative".

Data: We collected a total of 88 gaze sequences, one for each of the four tasks our 22 participants completed. The labeling on these gaze sequences produced an unbalanced data set with 540, 409 negative samples and 48, 033 positive samples. To create our training set, we extracted 76, 852 samples to create a balanced data set (containing 80% of the number of positive samples and the same number of negative samples). The test set was created by using the remaining 20% of the positive samples not included in the training set, as well as the same number of randomly-selected negative samples that were not included in the training set.

Prediction Model Details

Topology: We employed Lustig et al.'s Long-short Term Memory (LSTM) topology [17] in our approach considering the similarities between our problem statements. Lustig et al. successfully applied an LSTM [11] recurrent neural network –which can effectively identify patterns in large amounts of sequence data compared to rule-based approaches– to classify 2D gaze patterns of readers, while we attempted to classify gaze patterns of users navigating a virtual environment. Similar to their network structure, our network's outer layer had 512 hidden units of LSTM blocks. Furthermore, we used hyperbolic tangent as our output activation function and a hard sigmoid for our input (recurrent) activation function in these LSTM blocks. Finally, our outer layer consisted of one output unit with a sigmoid activation function. A visualization of our network's topology is provided in the supplementary material.

Training: We optimized our network using Adam [12], a gradient-based stochastic optimizer, with binary cross-entropy as our loss function.

In order to select the best performing model, we extracted 10% of our training data to be used as validation data while training. This validation data was held-out and used to measure the accuracy of the model at each epoch. In other words it was unchanged throughout training.

The training data was shuffled before each epoch. It did not contain any sample from the validation or test sets. The optimizer iterated over the training data in batches of 100 samples. We trained our model for 500 epochs and selected the network that had the best accuracy (100%) on the validation data at the 237th epoch as our classifier. Figure 8 shows the validation and training accuracy achieved at each epoch. Beyond the 237th epoch the validation and training accuracies generally begin to drop; we halted training the LSTM at the 500th epoch.

Implementation: We used the Python library Keras with a TensorFlow GPU back-end to set up and train our network. We trained our network in a GPU-based high performance computing cluster node, which had two 4-core Intel(R) Xeon(R) CPUs with a processor base frequency of 2.40GHz and 32GB of RAM. We utilized two GeForce GTX 670 GPUs in our training. We trained our model for 500 epochs which took approximately 75 hours on the GPU node.

6 EVALUATION OF MODEL PREDICTIONS

The sigmoid function on the final layer of the LSTM produces a score ranging from 0 to 1. Using a decision threshold $\sigma \in [0, 1]$ and this score, we can classify samples as negative ($< \sigma$) or positive ($\geq \sigma$). In the evaluation of the LSTM, we set the threshold σ to be 0.5.

Our model can predict users' need of navigation aid with high accuracy as evidenced by the accuracy, precision, recall and F-score of our classifier shown in Table 1. We achieved a 99.94% accuracy on our 19, 214 sample test set. Please refer to the Data Processing section for more details about how we created this test set.

We show the confusion matrix as further evidence of our model's performance in Figure 9. The confusion matrix shows a comparison between the numbers of positive and negative samples in the test set that were predicted correctly.

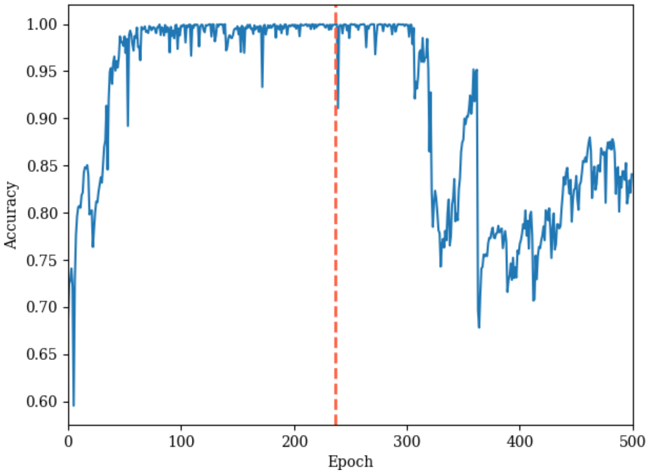


Figure 8: Validation accuracy of our model at every epoch of training. The accuracy was computed using the held-out validation data. The network corresponding to the 237th epoch was selected as our classifier.

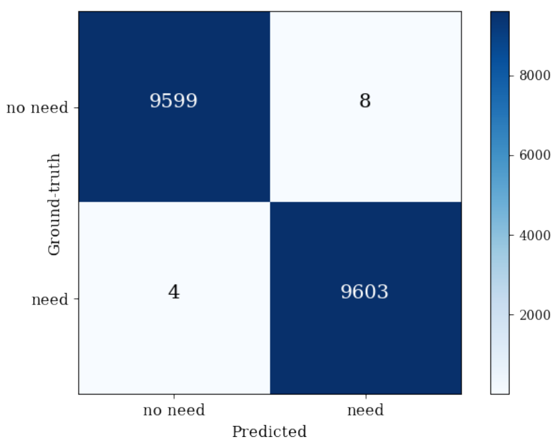


Figure 9: Confusion matrix showing the result of classifying the test set using our model, where "no need" and "need" denote prediction as negative (no need for navigation aid) and positive (needing navigation aid) respectively.

Accuracy	Precision	Recall	F-score
99.94%	99.92%	99.96%	99.94%

Table 1: Classification metrics on our test set.

We observed that our model correctly predicted the existence of navigation aid need (positive samples) more than 99% of the time, while erroneously predicted that a user did not need navigational aid less than 1% of the time. Similarly, our model correctly predicted that a user did not need aid (negative samples) more than 99% of the time, while mistakenly predicted that he/she needed navigation aid less than 1% of the time.

7 ADAPTIVE NAVIGATION AID

Anticipating when users need navigation is useful in designing interactive navigation tools. Using our navigation aid classifier, we were able to determine if a user needed navigation aid with only 7 seconds (350 time frames) of gameplay using his/her gaze. After gathering more than 7 seconds of data, we can slide our 7-second window to continuously classify new gaze points. Once we classified a user's gaze patterns as needing navigation aid, we displayed a tool to guide him/her to the destination. In our experiments, we realize this tool in the form of an adaptive arrow or a mini-map which are commonly used.

Figure 10: Users can set the sensitivity of the adaptive navigation aid.

Designers and end users of virtual experiences can select the level of navigation help they want using the "aid sensitivity control bar" shown in Figure 10. Users can select how sensitive the adaptive aid is with regard to classifying gaze patterns as needing navigation aid. Each level of sensitivity chosen on the UI corresponds to the number of times per minute the gaze must be classified as positive before the aid is shown. For example, Figure 10 shows a user selection that displays the aid only if 3 or more windowed gaze sequences in a minute have been classified as positive–pertaining to someone who would need navigation aid. Each windowed gaze sequence refers to data collected from 7 seconds (350 time frames) of navigation. In our user study, we set the sensitivity to 1 (i.e., we display the adaptive aid anytime a participant's gaze indicates he/she needs navigation help).

8 USER STUDY

Our adaptive navigation aid was designed to enhance navigation in VR. We conducted a user study to validate the effectiveness of our adaptive aid for guiding users to their destinations fast and with little frustration. We also compared our adaptive aid with conventional permanent aids (arrows and mini-maps) with regards to user satisfaction and immersiveness.

Participants: Fifteen healthy college and graduate students whose ages ranged from $18 - 30$ were recruited to participate in this study, akin to [3, 16, 23]. Participants gave written consent to participate in the IRB-approved user study.

Setup: We used the virtual reality headset setup same as that of our data collection experiment.

Tasks: To conduct the user study, we used the *VR City* scene that had been employed in the data collection experiment with some minor modifications. Because the *VR City* layout was designed to be maze-like, and we did not intend to

Figure 11: The *VR City* scene modified for our user study. A tower was added as a landmark to the scene. Neighborhoods of *VR City* were given different colors to help participants deduce their locations in the scene.

frustrate participants with the wayfinding task, we opted to facilitate navigation in this scene by slightly modifying it. Figure 11 shows the modified layout. We highlighted certain regions of the scene using different colors to make it easy for participants to localize their positions in the scene.

As landmarks play an important role in navigation [2, 4, 13], we added a tower to the scene, which was tall enough to be viewed from any location in the scene.

We asked the participants to complete tasks similar to the tasks designed for our data collection experiment. In the first half of each task, the participant was assigned to a starting point and was asked to follow a path to a specific fruit. In the second half, he/she was asked to return to the starting point under an aid condition.

In our user study, we specified a total of 80 tasks, each one being a combination of a starting point, a fruit and an aid condition (Figure 11 shows the starting points and fruit locations). Our 5 aid conditions are defined as: no aid, permanent arrow, adaptive arrow, permanent mini-map and adaptive mini-map. Figure 15 shows screenshots of the aids used. Every participant in our study was exposed to all of our aid conditions, to allow for a within-subject comparison between aids.

We cycled through all of these 80 tasks giving each participant 5 distinct tasks, such that he/she completes one task in each of our 5 aid conditions. The participant may complete a task with the same fruit or same starting point twice, but never the same fruit and starting point pairing. For example, participant *A* was given task 1 which asks him/her to find the *apple* from starting point 1, using the adaptive arrow on his/her way back. In task 5, participant *A* may be dropped

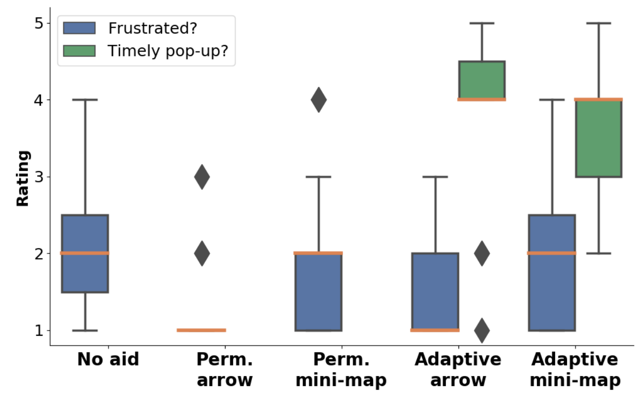


Figure 12: Participants of our user study answered the questions with a rating of 1 (strongly disagree) to 5 (strongly agree). The questions are listed in the user survey section.

in starting point 1 to find the *kiwi* using a permanent mini-map or dropped in starting point 3 to find the *apple* using a permanent min-map on the way back. Because we had 15 participants in total we were able to cycle through all of our tasks giving each participant a unique task. In the previous scenario, for example, other participants will not complete task 1 and task 5 since they were already assigned to participant *A*. This task design was created to prevent participants from being biased towards an aid condition due to the order it was given (e.g., the participant might find it easier to complete the task under the final condition, after becoming familiar with the scene in early tasks under other conditions) or its pairing with a starting point and fruit (e.g., participants might find it easy to return to starting point 3 from the *apple* because blue might be the most memorable color, and starting point 3 is in the blue neighborhood). To prevent the aid from biasing the participant's post-navigation survey decisions, he/she was not explicitly informed about which aid condition was present in the task prior to taking its corresponding survey.

User Survey: Participants were asked to complete a survey after each task to evaluate the performance of the aid given under the condition. We opted to use a 5-point Likert scale to evaluate the aids similar to [23]. The participants were asked to answer the following questions with a rating of 1 (strongly disagree) to 5 (strongly agree):

- You often feel frustrated during the navigation?
- The aid popped-up when you needed it?

Results

Adaptive Aid's Effects on Navigation. Figure 12 visualizes participants' answers to the question "*You often feel frustrated during the navigation?*" as a boxplot. A Friedman test revealed a significant effect of aid condition on frustration reported ($\chi^2(4) = 13.11$, $p = 0.01$). A post-hoc test using Wilcoxon

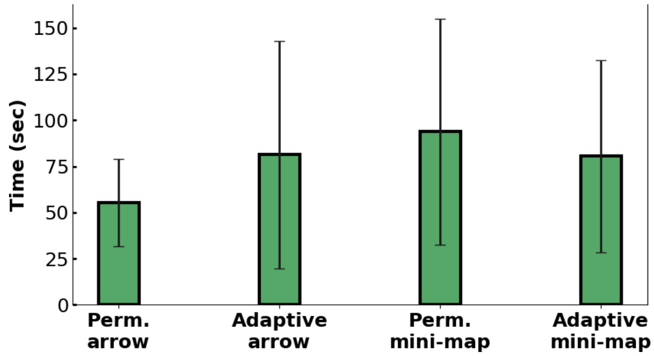


Figure 13: Average time taken by participants to complete our user study tasks.

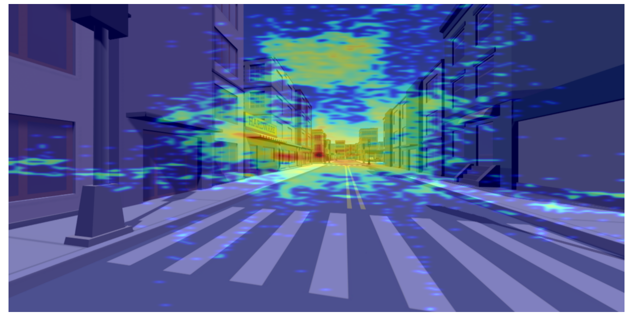


Figure 14: Participants' gaze when navigating without an aid in the user study. They looked at the top of the screen anticipating an aid to appear.

signed rank-test with Bonferroni correction showed that the median of navigation using our adaptive arrow (Md=1 strongly disagree) was reported to be less frustrating than navigation without an aid (Md=2 disagree) ($W = 11, p < 0.01$, $r = 0.33$). Figure 12 also shows a larger range of frustration level reported for users navigating with no aid, with some participants who reported a rating of 4 (agree). Comparatively, participants reported at most neutral frustration (rating of 3) after navigating using our adaptive arrow.

Our Wilcoxon signed rank-tests with Bonferrroni correction did not find any significant difference in the frustration reported between navigation without an aid (Md=2) and navigation with our adaptive mini-map (Md=2) ($W = 34.5$, $p = 0.14$). We did not find a significant difference in frustration reported between navigation without an aid (Md=2) and with a permanent mini-map (Md=2) ($W = 35, p = 0.15$). This could indicate that the mini-map is generally frustrating by nature as a navigation tool, contrasted with its arrow counterpart which received a lower frustration rating (Md=1, strongly disagree) when compared to navigation without an aid (Md=2) ($W = 0, p < 0.001, r = 0.4$).

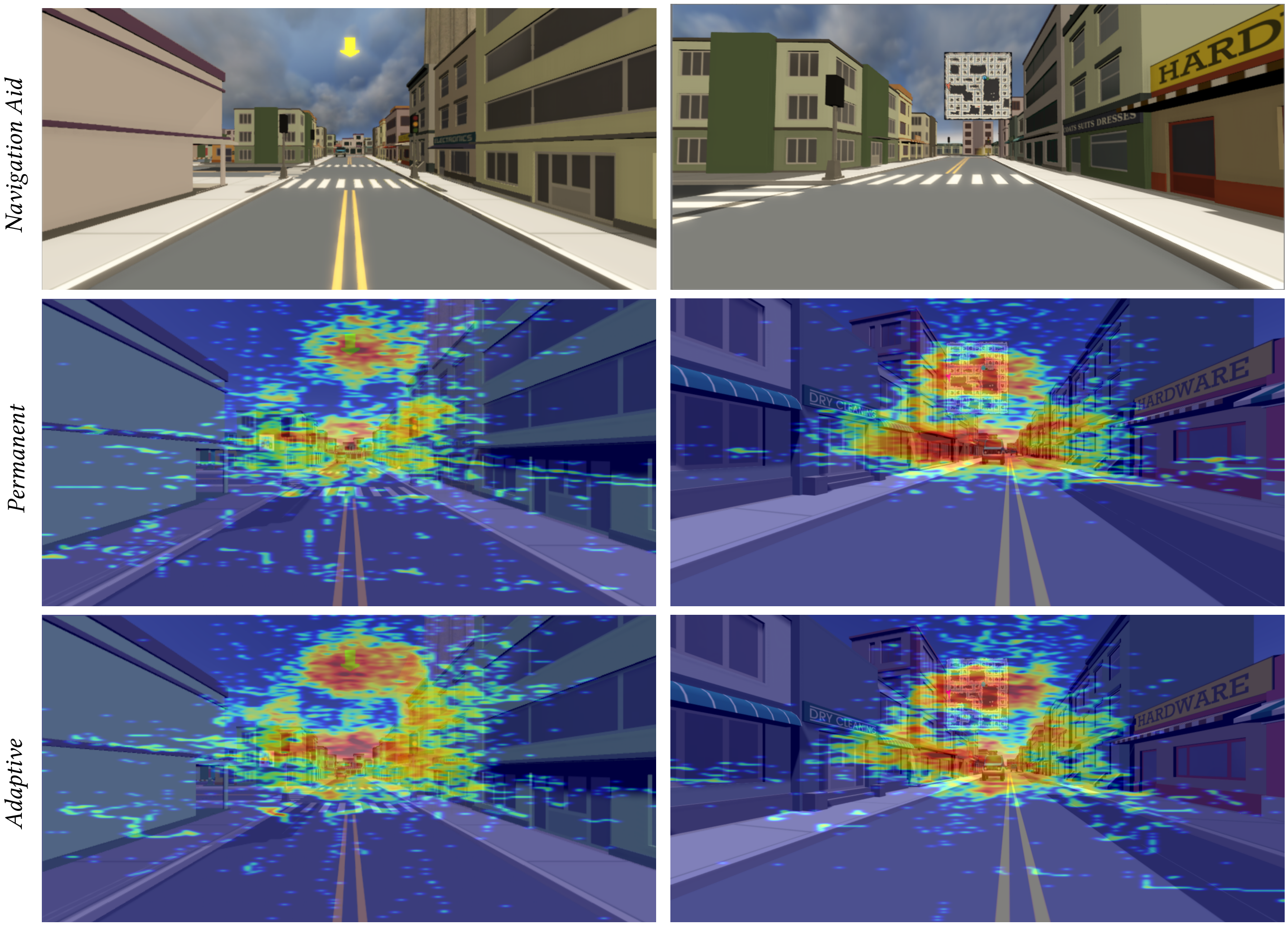
Comparing Adaptive Aids with Permanent Aids. Figure 12 showed that participants generally reported smaller frustration ratings after navigation with a permanent arrow (Md=1, strongly disagree) compared to our adaptive arrow (Md=1, strongly disagree) ($W = 0, p < 0.01, r = 0.39$). The majority of participants reported that they either disagreed (rating of 2) or strongly disagreed (rating of 1) that navigation with our adaptive arrow was frustrating. At most participants reported feeling neutral (rating of 3) frustration after navigating with our adaptive aid. This indicates that our adaptive arrow did not cause participants to feel frustrated, it did however increase frustration compared to permanent arrows. This suggests that there is a slight trade-off between the lowest possible level of frustration and immersiveness that the designer needs to consider while selecting a proper navigation aid to provide.

The post-hoc Wilcoxon rank-test with Boneferroni correction showed no significant difference between the frustration level reported by participants after navigating using the permanent map (Md=2) compared to the adaptive map (Md=2) ($W = 46, p = 0.43$). Similar to our comparison of navigation with no aid and adaptive mini-map, the lack of significance could be attributed to the abstruse nature of the mini-map. Showing the map when it was not necessary might have caused participants to second-guess themselves and to stop to look at the map, instead of relying on their intuition to where the destination might be.

Participants generally agreed that our adaptive aids appeared when needed, with our adaptive arrow receiving a score of (Md=4, agree) and our adaptive mini-map receiving a score of (M=4, agree) with regard to timely pop-up in Figure 12. We conducted a Wilcoxon signed-rank test with Bonferroni correction to compare the timeliness of our adaptive arrow and mini-map and found that the difference between these two aids followed a symmetric distribution around zero ($W = 39.5, p = 0.67$).

Figure 13 shows the average times taken by participants to complete our user study tasks. There was no significant difference in terms of task completion time when using the adaptive aids compared to when using their permanent counterparts. Participants completed the task in an average of (M=51.15, SD=23.65) seconds using the permanent arrow, not significantly different than and the average completion time (M=81.33, SD=61.68) of the participants while using the adaptive arrow ($W = 30.5$, $p = 0.09$). The Wilcoxon Signed-rank test with Bonferroni correction did not show a significant difference ($W = 55, p = 0.78$) in the average completion time of participants using the permanent ($M = 93.72$, SD=61.11) and adaptive mini-map ($M = 80.46$, SD=51.97) neither.

Gaze Patterns under Different Conditions. To visualize participants' visual attention under different conditions, we showed heatmaps of their average gaze points. Particularly, we analyzed the participants' gaze points inside and outside



(a) Arrow (b) Mini-map

Figure 15: Two types of aids used in our user study. (a) The arrow points the participant to the direction of the goal, (b) while the mini-map shows the participant's position and direction. Gaze heat-maps of participants traversing with each of these navigation aids are shown. Red and blue regions respectively indicate high and low amounts of visual attention.

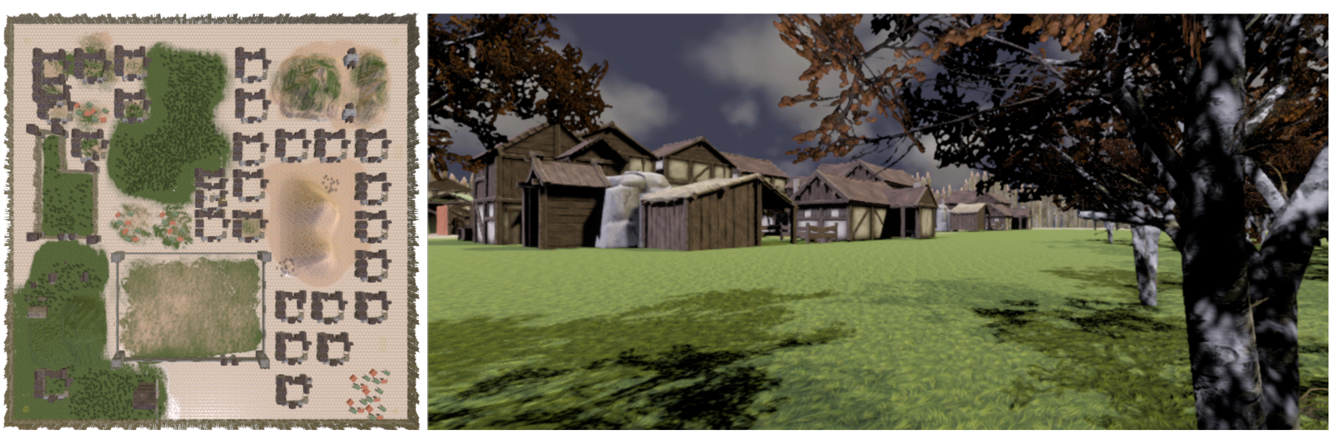


Figure 16: *VR Village*, a rural scene we created to evaluate our classifier's accuracy.

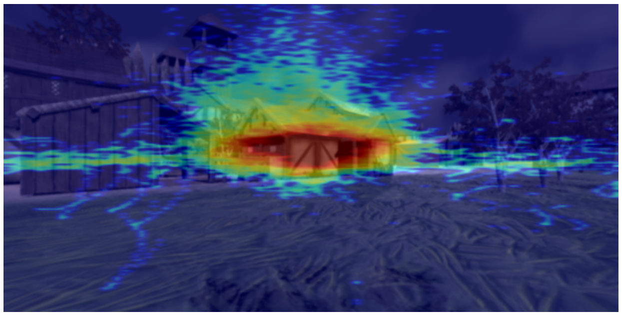


Figure 17: Gaze heatmap of participants navigating *VR Village*

the aid's region under different aid conditions. The arrow and mini-map aid regions are shown in the supplementary material.

We found that participants looked at the top of the screen even when no aid was shown as depicted in Figure 14. It is plausible that some participants were anticipating an aid to be shown to them after completing one or more tasks with a navigation aid.

Arrows: The adaptive arrow caused participants to explore a larger portion of the screen outside of the aid's region. Using a Wilcoxon Signed- rank test ($W = 18, p = 0.02, r = 0.62$), the region outside of the *permanent arrow* received an average gaze fixation duration of (M=43.78, SD=16.21) seconds, significantly (27.04 seconds) shorter than the region outside of the *adaptive arrow* has received (M=70.82, SD=41.61).

It is obviously true that using the permanent arrow is less frustrating and more efficient than the adaptive arrow. However, when comparing the gaze patterns in Figure 15, it is clear that the adaptive arrow surpassed the permanent arrow in keeping participant's attention on the scene as opposed to on the aid.

To conclude, our adaptive arrow is effective in situations which the designer *prefers players to look around the virtual environment, without causing them to feel as frustrated as navigating without an aid*, or significantly increasing the task completion time.

Mini-maps: The adaptive mini-map did not significantly change the participants' gaze behavior compared to the permanent mini-map. The gaze region outside of the mini-map received an average gaze fixation duration of (M=70.37, SD=50.5) and (M=52.4, SD=20.3) seconds in the case of permanent and adaptive mini-maps respectively. It appears that participants explored the bottom portion of the screen more when using the adaptive mini-map than when using the permanent mini-map. However, this contrast was not found to be significant as in the case of the arrows ($W = 51, p = 0.61$).

The mini-map is inherently less usable than the arrow. It is possible that participants felt frustrated more often whilst navigating with the adaptive min-map, causing it to be triggered to them frequently, thus diverting their gaze away from the outside of the aid region. Conversely, the adaptive arrow gave participants the opportunity to utilize their instinctive cognitive navigation skills without adding a layer of confusion. Overall, we believe that a participant's own wayfinding skills combined with an intuitive adaptive navigation aid like the adaptive arrow might have been ideal in completing a navigation task.

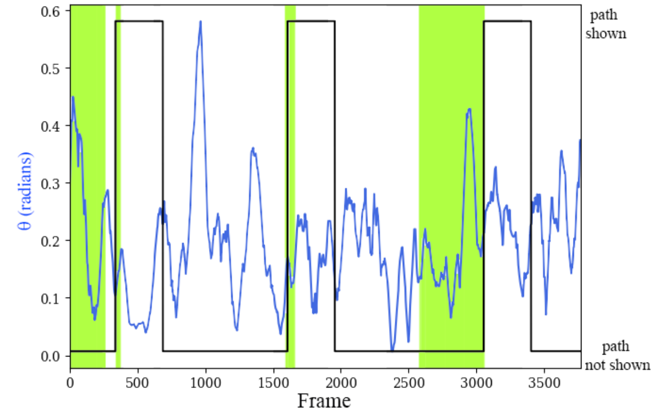


Figure 18: Relationship between a participant's gaze, aid requests and our classifier's prediction in *VR Village*. Green shows regions when our classifier predicted the navigation aid need just before the participant requested help.

9 DISCUSSION

Preliminary Test with a Different Scene: To test our classifier's performance on virtual environments the LSTM was not exposed to, we created a new scene *VR Village* (shown in Figure 16). Using the same data collection experiment setup in Section 4, We asked 12 participants, with ages ranging from $18 - 29$, to complete the same data collection tasks by following a path to a fruit and returning to the starting point. This data set was treated as an additional test set, and the classifier trained with the *VR City* data set was able to predict participant's need for navigation aid with a 80.68% accuracy.

We suspect that the drop in accuracy was due to the change in gaze behavior of participants navigating this rural scene. This is evident in Figure 17 showing the gaze heat-map of participants navigating *VR Village*. Participants viewed *VR Village* in an oval pattern with more spread around the horizon, while they viewed *VR City* in a U-shaped pattern (see Figure 14) focusing on the city blocks to localize their position. Unlike in the urban *VR City* scene where participants could use buildings, signs and counting street blocks to find their way; in the rural *VR Village*, participants relied on mountains, hills and trees to localize their position.

For some participants, the change in gaze patterns was minor and we were able to anticipate when they needed navigation aid. An example is shown in Figure 18. However, in order to maintain our high accuracy for all users it seems preferable to detect the type of the scene and task users are navigating and employ a navigation aid need classifier that has been trained on scenes and tasks of a similar type. We could also request that users press the hint button when completing a new task in a new scene. Afterward, we could use this data to fine-tune our mode.

Applications: Adding immersion to the virtual experience is not the only benefit of our adaptive aid. Our aid could be used by level designers to unintrusively collect data from users about the virtual experience. This could be helpful in allowing the designers to analyze the navigation difficulty of environments under design.

Moreover, our classifier and sensitivity bar mentioned in Section 7 can allow users to adjust the navigation difficulty according to their wayfinding skill level. Also, if level designers prefer users to explore the environment and not rely on a navigation aid, they can start users with a highly sensitive adaptive aid, but gradually decrease the sensitivity as users become more accustomed to the navigation experience.

Limitations and Future Work: Although eye-tracking embedded VR headsets are promising, they still do not attain the precision of standard eye-trackers. Faulty calibration in the FOVE resulted in several noisy sessions, in addition to being less comfortable compared to the more lightweight traditional eye-trackers. As the technology behind eye-tracking headsets matures, these limitations will hopefully subside.

Furthermore, IR-based VR trackers do not generalize well across all eye and skin colors. This phenomenon was explored by Feit et al. [5] using the IR-based tracker TobiiEye X. Li et al. [15] noted varying cross-user accuracy in the FOVE due to differences in amount of reflected light across skin colors and pupil light absorption across eye colors.

The varying latencies, frequencies and accuracies of head-mounted displays are likely to affect our prediction model. In other words, a head-mounted display with different specifications than the FOVE, which can not be remedied by adjusting the sampling rate, is likely to require fine-tuning of our model.

We found it challenging to label our dataset with the definite moments that our subjects felt a need for a navigation aid. From our observations on the subjects' behavior and several labeling attempts in preliminary experiments, we estimated the time that the subjects might have needed the aid before they pressed the hint request button. This estimation could have been improved by using an additional tool to serve as a reference point for the cognitive state subjects are in when needing help. For example, an Electroencephalography (EEG) sensor could have been utilized for monitoring subject's brain activity, and any discernible recordings before triggering the hint button could have been used as markers.

Although we extensively explored our adaptive tool's effectiveness in our user study, we have yet to conduct an extensive usability study. The NASA TLX (Task Load Index) [9] could help us identify ways to improve users' experience with our tool. We could also study the usability of a hybrid solution: an application can employ our adaptive navigation aid and allow users to manually trigger the aid when needed as well. We could also adaptively fade out the aid when users no longer need the help. These hybrid solutions could possibly achieve the permanent arrow's low frustration level and the adaptive arrow's immersiveness. We empirically showed that participants found navigation using the permanent arrow less frustrating, while it pulls their gaze away from the scene and puts it onto the aid. A best-of-both-worlds solution could maintain the adaptive arrow's immersiveness and the permanent arrow's low-level of frustration.

Given the reasonable accuracy of our classifier, it could be extended to serve other navigation-related purposes. For example, the classifier's predictions could be used by level designers for analyzing the navigability of a virtual environment in design. If a specific region of the environment frustrated a significant number of users, our prediction model could be applied to detect that and visualize it for the level designer.

Finally, we explored the use of gaze patterns for facilitating virtual environment navigation only. It would be interesting to investigate whether our approach can be applied in a real-world setting, for example, popping up navigation hints through augmented reality glasses to assist a user in navigating an unfamiliar environment in the real world.

10 CONCLUSION

In this paper, we investigated the use of gaze patterns for classifying the need for navigation aid in virtual reality. We devised a model that is able to predict when users needed navigation help and used the predictions to adaptively aid them in finding the right way. We validated the efficacy of our adaptive navigation aid through a user evaluation study, and showed its potential in improving the engagement of users in virtual navigation while effectively guiding them to their destinations.

11 ACKNOWLEDGMENTS

We thank Kristen Laird for her help in conducting the user studies. This research is supported by the National Science Foundation under award number 1565978. We are grateful to the anonymous reviewers for their constructive comments.

REFERENCES

[1] Michael Burch. 2017. Which Symbols, Features, and Regions are Visually Attended in Metro Maps? In *Intelligent Decision Technologies 2017*. Springer International Publishing, 237–246.

[2] Chin-Kai Chang, C Siagian, and L Itti. 2010. Mobile Robot Vision Navigation & Localization using Gist and Saliency. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE.

[3] Lung-Pan Cheng, Sebastian Marwecki, and Patrick Baudisch. 2017. Mutual Human Actuation. In *UIST*. ACM, New York, NY, USA, 797–805.

[4] Rudolph Darken and Barry Peterson. 2001. Spatial Orientation, Wayfinding, and Representation. (2001).

[5] Anna Maria Feit, Shane Williams, Arturo Toledo, Ann Paradiso, Harish Kulkarni, Shaun Kane, and Meredith Ringel Morris. 2017. Toward everyday gaze input: Accuracy and precision of eye tracking and implications for design. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 1118–1130.

[6] Jonathan Gandrud and Victoria Interrante. 2016. Predicting Destination Using Head Orientation and Gaze Direction During Locomotion in VR. In *Proceedings of the ACM Symposium on Applied Perception (SAP '16)*. ACM, New York, NY, USA, 31–38.

[7] Amit Garg, Joshua A. Fisher, Wesley Wang, and Karan Pratap Singh. 2017. ARES: An Application of Impossible Spaces for Natural Locomotion in VR. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. ACM, New York, NY, USA, 218–221.

[8] Jack Goldfeather. 2012. Adaptive Redirected Walking in a Virtual World. In *2012 IEEE VR Workshop on Perceptual Illusions in Virtual Environments*. IEEE.

[9] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[10] Rorik Henrikson, Bruno Araujo, Fanny Chevalier, Karan Singh, and Ravin Balakrishnan. 2016. Multi-Device Storyboards for Cinematic Narratives in VR. In *UIST*. ACM, New York, NY, USA, 787–796.

[11] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9, 8 (1997), 1735–1780.

[12] Diederik P Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980* (2014).

[13] Christian Lander, Frederik Wiehr, Nico Herbig, Antonio Krüger, and Markus Löchtefeld. 2017. Inferring Landmarks for Pedestrian Navigation from Mobile Eye-Tracking Data and Google Street View. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. ACM, New York, NY, USA, 2721–2729.

[14] Jooyeon Lee, Manri Cheon, Seong-Eun Moon, and Jong-Seok Lee. 2016. Peripersonal Space in Virtual Reality: Navigating 3D Space with Different Perspectives. In *UIST (UIST '16 Adjunct)*. ACM, New York, NY, USA, 207–208.

[15] Tianxing Li, Qiang Liu, and Xia Zhou. 2017. Ultra-Low Power Gaze Tracking for Virtual Reality. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*. ACM, 25.

[16] Pedro Lopes, Alexandra Ion, and Patrick Baudisch. 2015. Impacto: Simulating Physical Impact by Combining Tactile Stimulation with Electrical Muscle Stimulation. In *UIST*. ACM, New York, NY, USA, 11–19.

[17] Joakim Lustig. 2016. Identifying Dyslectic Gaze Pattern: Comparison of Methods for Identifying Dyslectic Readers based on Eye Movement Patterns.

[18] Andreas Möller, Matthias Kranz, Stefan Diewald, Luis Roalter, Robert Huitl, Tobias Stockinger, Marion Koelle, and Patrick A. Lindemann. 2014. Experimental Evaluation of User Interfaces for Visual Indoor Navigation. In *ACM SIGCHI*. ACM, New York, NY, USA, 3607–3616.

[19] Roberto A. Montano Murillo, Elia Gatti, Miguel Oliver Segovia, Marianna Obrist, Jose P. Molina Masso, and Diego Martinez Plasencia. 2017. NaviFields: Relevance Fields for Adaptive VR Navigation. In *UIST*. ACM, New York, NY, USA, 747–758.

[20] Cuong Nguyen, Stephen DiVerdi, Aaron Hertzmann, and Feng Liu. 2017. CollaVR: Collaborative In-Headset Review for VR Video. In *UIST*. ACM, New York, NY, USA, 267–277.

[21] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L. Davidson, Sameh Khamis, Mingsong Dou, Vladimir Tankovich, Charles Loop, Qin Cai, Philip A. Chou, Sarah Mennicken, Julien Valentin, Vivek Pradeep, Shenlong Wang, Sing Bing Kang, Pushmeet Kohli, Yuliya Lutchyn, Cem Keskin, and Shahram Izadi. 2016. Holoportation: Virtual 3D Teleportation in Real-time. In *UIST*. ACM, New York, NY, USA, 741–754.

[22] Yun Suen Pai, Benjamin Outram, Noriyasu Vontin, and Kai Kunze. 2016. Transparent Reality: Using Eye Gaze Focus Depth As Interaction Modality. In *UIST (UIST '16 Adjunct)*. ACM, New York, NY, USA, 171–172.

[23] Ken Pfeuffer and Hans Gellersen. 2016. Gaze and touch interaction on tablets. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM, 301–311.

[24] L. Piccardi, M. De Luca, R. Nori, L. Palermo, F. Iachini, and C. Guariglia. 2016. Navigational Style Influences Eye Movement Pattern during Exploration and Learning of an Environmental Map. *Front Behav Neurosci* 10 (2016), 140.

[25] Sharif Razzaque, Zachariah Kohn, and Mary C. Whitton. 2001. Redirected Walking. In *Eurographics 2001 - Short Presentations*. Eurographics Association.

[26] Deba Pratim Saha, R. Benjamin Knapp, and Thomas L. Martin. 2017. Affective Feedback in a Virtual Reality Based Intelligent Supermarket. In *UbiComp*. ACM, New York, NY, USA, 646–653.

[27] Charles Lima Sanches, Koichi Kise, and Olivier Augereau. 2017. Japanese Reading Objective Understanding Estimation by Eye Gaze Analysis. In *UbiComp*. ACM, New York, NY, USA, 121–124.

[28] Junichi Shimizu and George Chernyshov. 2016. Eye Movement Interactions in Google Cardboard Using a Low Cost EOG Setup. In *UbiComp*. ACM, New York, NY, USA, 1773–1776.

[29] S. J. Tetewsky and C. J. Duffy. 1999. Visual Loss and Getting Lost in Alzheimer's Disease. *Neurology* 52, 5 (mar 1999), 958–958.

[30] Jayson Turner. 2013. Cross-device Eye-based Interaction. In *UIST (UIST '13 Adjunct)*. ACM, New York, NY, USA, 37–40.

[31] Eduardo Velloso and Marcus Carter. 2016. The Emergence of EyePlay: A Survey of Eye Interaction in Games. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '16)*. ACM, New York, NY, USA, 171–185.

[32] K. Vlcek and J. Laczo. 2014. Neural correlates of Spatial Navigation changes in Mild Cognitive Impairment and Alzheimer's Disease. *Front Behav Neurosci* 8 (2014), 89.