Eye Tracking-based LSTM for Locomotion Prediction in VR

Niklas Stein*†
Institute for Psychology,
University of Muenster

Gianni Bremer*‡
Institute for Psychology,
University of Muenster

Markus Lappe§
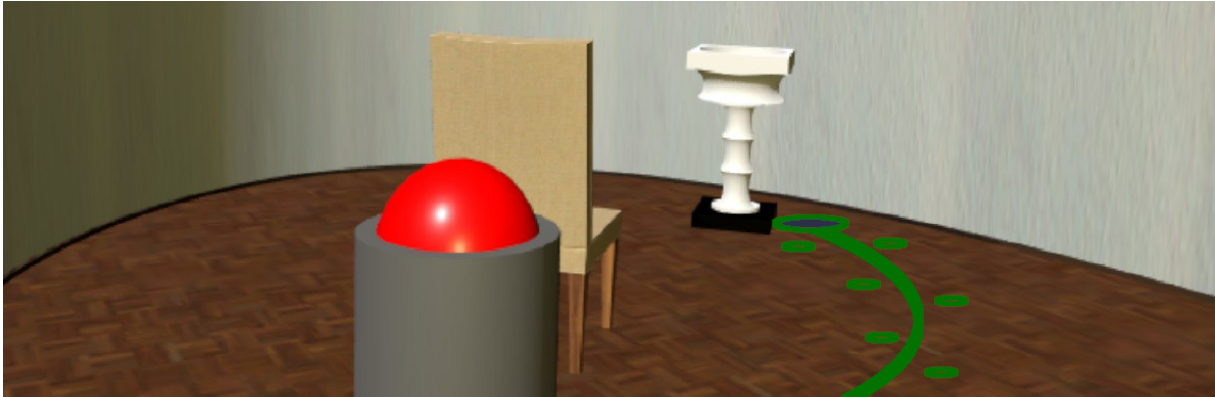Institute for Psychology,
University of Muenster

Figure 1: User perspective of the experiment while walking to a target and avoiding an obstacle.

Abstract

Virtual Reality (VR) allows users to perform natural movements such as hand movements, turning the head and natural walking in virtual environments. While such movements enable seamless natural interaction, they come with the need for a large tracking space, particularly in the case of walking. To optimise use of the available physical space, prediction models for upcoming behavior are helpful. In this study, we examined whether a user's eye movements tracked by current VR hardware can improve such predictions. Eighteen participants walked through a virtual environment while performing different tasks, including walking in curved paths, avoiding or approaching objects, and conducting a search. The recorded position, orientation, and eye-tracking features from 2.5 s segments of the data were used to train an LSTM model to predict the user's position 2.5 s into the future. We found that future positions can be predicted with an average error of 65 cm. The benefit of eye movement data depended on the task and environment. In particular, situations with changes in walking speed benefited from the inclusion of eye data. We conclude that a model utilizing eye tracking data can improve VR applications in which path predictions are helpful.

Keywords: Virtual Reality, Eye Tracking, Locomotion, LSTM, Path Prediction, Machine Learning, Gaze

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual reality; Applied computing—Law, social and behavioral sciences—Psychology—; Computing methodologies—Machine learning—Machine learning approaches—Neural networks;

---

*Niklas Stein and Gianni Bremer are co-first authors.
†Niklas.Stein@uni-muenster
‡Gianni.Bremer@uni-muenster.de
§mlappe@uni-muenster.de

1 Introduction

Walking is the most natural and most immersive way of moving through a virtual environment (VE) [47]. It enhances presence and allows users to unconsciously acquire spatial knowledge [33, 52]. However, walking creates a set of physical problems for virtual reality (VR) applications. Collisions with physical walls, objects or other users in real space need to be prevented at all times. Therefore, large VEs either require large tracking areas or methods such as redirected walking (RDW). With methods that estimate intentions or predict future actions of walking users, RDW and other applications such as programmable interaction patterns of avatars, automatic, user behavior based level design could become more effective.

RDW is the subtle and unbeknownst redirection of the user to change her trajectory in real space while keeping her perception of movement in virtual space the same. RDW is limited by the user's manipulation threshold [46]. If visually presented and physical walking paths diverge too much, the manipulation is noticed by the user. Although the detection threshold can be slowly adapted over time through learning, the general limitation remains [8].

The implementation of RDW requires applying these manipulations automatically using RDW controllers. Scripted steering controllers have a predefined set of rules based on given information about the VE and the physical space [42]. As a result, RDW manipulations are applied whenever a participant reaches a predefined position. Accordingly, this type of controller needs to be carefully readjusted whenever the VE or the available physical space changes. In the past, several methods to automatize this readjustment process have been developed [e.g. 59]. The available virtual walking paths can generated automatically from the environmental data [58]. Equipped with predefined probability scores for each path and a given skeleton map of the virtual and physical space, the controller then chooses the best manipulation from a predefined set during walking. Generalized controllers follow a different approach. Instead of using information about the VE, algorithms such as steer-to-center, steer-to-orbit, steer to way-points, or steer in figure-eight patterns are designed to work in any VE by steering users along specified physical paths [40].

RDW controllers can greatly benefit from predictions or assumptions about the upcoming behavior of the user. Given predictions of the user's virtual path, manipulations can be applied earlier and with

a smaller divergence between real and virtual paths. One source for the prediction in unknown environments can be recurring behavior patterns. For example, assumptions such as typical walking paths [25] have proven useful in this context. Besides general assumptions, predictions of upcoming locomotor activities may also be gained from observing the user's prior behavior and actions, such as the trajectory of the user's path over the immediate past and the orientation of her head and body during that time.

Another potential source of information is the user's gaze pattern. When walking, humans typically use specific gaze strategies to inspect their future path, monitor their next target, and avoid obstacles [18, 22, 49]. Therefore, the pattern of gaze during walking contains valuable information about the user's locomotor intentions. Gaze tracking has recently been incorporated into commercially available head mounted displays (HMD). Although their current quality does not match that of research-grade eye trackers [34, 43, 45] it appears still possible to use gaze information to improve locomotor predictions of VR users [6, 16, 56, 57]. Since gaze behavior depends to some degree on the task (e.g. obstacle avoidance is different from walking straight to a target) the usefulness of gaze data for walking prediction might likewise depend on the situation.

In the present study, we explore how combinations of position, orientation, and gaze data can be used to predict a user's future position during walking in VR, and how the prediction quality depends on the different data features and task demands. Our prediction model uses an artificial neural network for time series prediction, the Long short-term memory (LSTM) architecture [19], which can be fitted to the relevant data without prior assumptions and has been used successfully, for example, for prediction of public pedestrian traffic [e.g. 2] as well as in RDW [10].

2 RELATED PREVIOUS WORK

2.1 Gaze Behavior During Walking

Eye movements have a tight correlation with other motor action because we need to move our eye to targets of interest to collect the visual information we need for good action control [31]. Usually eye movements to a behavioral target precede any other motor action [18, 32]. Therefore, they present a rich signal for the estimation of action intention [3, 6, 16, 56, 57]. However, during walking this does not mean that people lock their gaze on a future target, for example a door, at all times. Instead, during walking, eye movements serve the identification of both targets and obstacles. For example, walkers often look at the ground in front a few steps ahead for safe placements of the feet, particularly in uneven terrain [9, 20, 21, 36, 49]. This involves not only a direction of gaze towards the ground but also a pitch of the head [35]. Before approaching a goal, however, walkers typically direct their gaze towards the goal [13, 22].

Eye movements are also linked to changes of direction during walking [22]. When walking in a curve, for example, walkers typically direct their gaze inward from the curve [17, 26]. Furthermore, eye movements are involved in deciding between alternative targets [53, 56] and in searching of targets between distractors [29]. Thus, eye movements during walking depend on task demands [50]. Therefore, during natural behavior, eye movements might provide a useful, though not straightforward, signal for a user's intention and future action that might be extracted in deep learning approaches for action prediction. Their usefulness is potentially different for different tasks. For the present investigation we chose three tasks to cover a set of general locomotor scenarios in which eye movements likely play a role: searching for a target amongst distractors in a room, walking along a curved path, and avoiding an obstacle.

2.2 Locomotor Prediction

Different methods for the prediction of future trajectories of users in VR have been proposed in the past. One approach relies on the analysis of possible or probable paths in the current environment

[58, 59]. Available virtual walking paths are generated automatically from the environmental data and probability scores for each path along with a skeleton map of the virtual and physical space allow, for example, to choose an optimal manipulation for RDW. This approach, however, is tied to the specific environment and needs environmental information.

A different approach focuses on observation of the user's prior actions to predict the future trajectory. Zank and Kunz [56] used eye-tracking data to predict the next locomotion target. In their experiment, two predefined target positions were presented and the user was either instructed to go to one of the targets or freely choose one of them. Based on the chosen targets, they also compared different models that use previous movements to calculate probabilities for the two targets based on assumptions about human walking behavior [1, 15, 57] and graph representations of the environment [38]. In particular, in a narrow T-shaped corridor with little open space, models using eye data were able to provide accurate predictions earlier than models without eye data. Later in a trial and in cases with open space, the overall prediction accuracy was higher and including eye data had no benefit for the prediction at those times.

Gandrud and Interrante [16] also implemented a binary choice between two walking targets. Users walked along a virtual hallway at the end of which they had to decide between two targets to approach. In the experiment they measured head direction, gaze direction and the position relative to the virtual hallway midline and compared these three measures for predicting the chosen target. They found that both head orientation and gaze orientation had the potential to be useful in predicting a person's future direction of locomotion.

The use of eye data in these approaches assumes that the gaze position directly precedes the direction of human walking. Indeed, there is evidence supporting this notion [7, 22, 31, 51]. However, since eye movements during walking also depend on task demands [50] a more complex processing procedure of gaze data could allow an even better prediction. Moreover, these pioneering previous studies only distinguished between binary walking decisions. Further scenarios with less restrictions need to be evaluated to advance the use of behavioral measures for locomotor predictions.

A promising current approach to predict future positions is the use of deep learning models. These models can be fitted to the relevant data without prior assumptions. They have been successful, for example, with respect to public pedestrian traffic [e.g. 2, 55] or for the prediction of future gaze directions [12, 14, 23, 24, 54]. With respect to walking prediction, Cho et al. [10] presented a preliminary study in which they implemented a deep learning model for locomotion prediction in the context of RDW. They used head position and orientation to train an LSTM model [19] to predict the user's position 100 frames (about 1 second) into the future while the user navigated a maze. They report that the prediction worked well for two example users. However, their model is limited to the pre-defined maze map they used and did not include gaze data.

3 SCOPE OF THIS STUDY

In the present work, we created a scenario in which participants fulfilled a range of typical VR tasks, including walking in curved paths, avoiding or approaching objects, and conducting a search. Based on the walking and gaze data, we trained and evaluated different indoor path prediction models with an LSTM architecture that had no information about the used VE. We were especially interested in whether a user's eye movements during walking could be a useful addition to the model. Therefore, we analyzed the stability of the eye tracking over the duration of the study, described typical data patterns during different tasks and evaluated under which circumstances eye tracking data contributes to a better prediction of future position.

4 METHODS

4.1 Participants

Eighteen participants (8 female) completed the experiment. The participants' age ranged from 20 to 47 years ($\mu = 27, \sigma = 6.34$). Participants gave informed written consent and the experimental procedures were approved by the Ethics Committee of the Department of Psychology and Sports Science of the University of Muenster. Apart from the two authors who participated in the experiment (N.S. and G.B.) participants were naïve to the purpose of the experiment. Two additional participants we initially acquired had to be excluded due to failed recordings.

4.2 Materials

The virtual environment was displayed in an HTC Vive Pro Eye with a resolution of 1440×1600 pixels per eye at a frame rate of 90 Hz and a nominal field of view of 110 degrees. The experiment was conducted at the VR laboratory of the Department of Psychology and Sports Science of the University of Muenster. An area of 6 x 11 m was tracked using 6 Vive Lighthouses 2.0. During the experiment, all positional tracking data were Kalman filtered. The experiment was run using Unity3D on an MSI GE63VR 7RF Raider notebook equipped with an NVIDIA GTX 1070 graphics card. The notebook was carried in a backpack and supplied with power by a cable from the ceiling. A Vive tracker was attached to the backpack to record the body orientation and a Vive controller was used as input device. Eye tracking data were recorded using the integrated Tobii eye tracker of the Vive Pro Eye with a nominal accuracy of 0.5–1.1 degrees within a field of view of 20 degrees at an output frequency of 120 Hz and a trackable field of view of 110 degrees.

4.3 Eye Tracking Procedure and Questionnaires

At the beginning of the experiment, eye tracking calibration was done using the calibration method provided by Tobii Software.

In the last decade, eye tracking cameras have been included in several commercially available VR headsets [11]. Different aspects like latency, accuracy and precision of eye tracking data in VR have been evaluated [34]. Under optimal conditions, the HTC Vive Pro Eye showed eye tracking data delays of around 50 ms and an accuracy below 1 degree in central, but up to 10 degrees in very peripheral positions (27 degrees). Eye tracking precision varied from 1.4 to 3.5 degrees [43, 45]. This makes the Vive Pro Eye suitable for a rough online estimation of gaze in VR, although it is not clear how side effects such as slippage through head movements during natural walking affect the measurements. To evaluate possible slippage of the HMD on the head and therefore the eye tracking calibration stability before and after the experiment, 16 participants did an additional, custom-made, simple calibration procedure with 5 fixation positions before and after the experiment.

The simulator sickness questionnaire [27] (SSQ) was completed before and after each session. Additionally, the participants completed the Slater-Usoh-Steed questionnaire for immersion [44] (SUS) after the session. Both questionnaires were translated into German. The authors did not participate in the questionnaires.

4.4 Tasks and Virtual Rooms

The virtual environment was divided into three rooms: Search Room (for an example see Fig. 4a in section 5), Transition Corridor (example in Fig. 4b) and Obstacle Room (for an example see Fig. 1) in which all participants did ten trials including different tasks. To provoke a lot of natural walking in a limited physical space, Search Room and Obstacle Room were mapped onto the same physical space in an impossible room scenario [48]. Whenever a participant moved through the Transition Corridor to the door on the other side, an entry to the room opened on the other side and the interior changed. Participants were asked to maintain a natural walking speed during the experiment while performing the following instructed tasks.

4.4.1 Search Room - Free Exploration

In the Search Room, participants had to look for a search object among six identical looking distractor objects. One object was placed in the center and the others formed a hexagon around it (see Figure 4a). All objects had a random yaw direction orientation and a distance of 2 m to their next neighbour objects. Whenever an object was reached, the participants could test whether this object was the search object by holding the controller close to it while pressing the trigger on the back of the controller. The result was signaled by a red or green light above the object. In addition, a sound was played if the search object had been found.

During the task, the participants were free to decide which object they wanted to go to next and did not know which object was the search object beforehand. The target position was individually pseudo-randomized for each participant. After completing the task, participants walked through the Transition Corridor to the Obstacle Room.

4.4.2 Transition Corridor - Curved Path

The Transition Corridor connected the two other rooms (see Figure 4b) and followed a curved path with a radius of 5.5 m. Participants had to walk through the corridor to pass between the rooms. Data from the Transition Corridor was obtained to investigate walking along a curved path. Since the participants went back and forth between the rooms, ten right curves (Search Room to Obstacle Room) and nine left curves (Obstacle Room to Search Room) were recorded for each participant.

4.4.3 Obstacle Room - Straight Path and Obstacle Avoidance

In the Obstacle Room, participants were instructed to walk to a target object while avoiding an obstacle that might be positioned along the way. For each 4 m walk, the participants first positioned themselves on a white field in front of a pole with a red button. Pushing the button with the controller made the pole and button disappear, and the target object and the obstacle (a chair) appear. The task had 4 different conditions: obstacle centered, obstacle 30 cm to the left, obstacle 30 cm to the right and no obstacle. During each visit to the Obstacle Room, these 4 conditions were run in pseudo-randomized order. Figure 1 shows the button, the obstacle and the target at example positions and a typical walking path with footsteps. Note that in the real experiment, the three objects were never visible at the same time. The obstacle was placed in the middle between the button and the target. After selecting the target with the controller (by pressing the trigger at the back while holding it close to the target), target and obstacle disappeared and a new red button appeared at a new location in the room. The participants then went to the new position and repeated the procedure. After four walks in each trial, the participants returned to the Search Room via the Transition Corridor.

4.5 Data

All raw data files are freely available from https://osf.io/b43uv/. For the analysis, data were sampled at intervals of 50 ms. Periods with missing data and periods in which the participant remained standing without any locomotion (threshold = 0.15 m/s) were removed. Such periods of prolonged standing occurred, for example, when the participants did not immediately start walking at the beginning of the experiment. Naturally, the eye tracker missed data whenever the participant briefly blinked. In this case, missing eye-tracking values were linearly extrapolated from the data before the blink.

For the prediction model we distinguished three categories of data:

1. Positional data from the Vive's infrared tracking system.

2. Orientation angles from the inertial measuring units (IMU).

3. Eye tracking data from the Vive Pro Eye's eye tracking system.

4.6 Prediction Model

4.6.1 Features and Labels

The model aimed to predict the user's location in the VR environment 2.5 s into the future. Thus, the output of our prediction model was defined as the direction vector from the current head position (X_t^H, Y_t^H, Z_t^H) to the position 2.5 s into the future $(X_{t+2.5s}^H, Y_{t+2.5s}^H, Z_{t+2.5s}^H)$.

Since our prediction model should be based on user behavior without information about the environment, it needed to use a coordinate system in a reference-frame attached to the user and not to the environment. For the present study, we used a head-fixed coordinate system, which has shown the best results previously (details in [6]).

To set up this coordinate system, we used the average head orientation in the horizontal plane of the input sequence $(\overline{\Phi_{t-i}^H}, \overline{\Theta_{t-i}^H}, \overline{\Psi_{t-i}^H})$ to create a fixed reference yaw angle that was used to describe each input-output-pair. This reference angle, along with its orthogonal directions in the horizontal and vertical planes, provided the axes of the head-fixed coordinate system. Thus, the label direction vectors $(\vec{F_t})$ were rotated using the reference yaw angle. Lower case letters are used to express variables in the new coordinate systems. (e.g. f, ψ, θ).

$$\vec{F_t} = (F_t^X, F_t^Z) = (X_{t+2.5s}^H - X_t^H, Z_{t+2.5s}^H - Z_t^H)$$
$$f_t^x = \cos(-\overline{\Psi_{t-i}^H})F_t^X - \sin(-\overline{\Psi_{t-i}^H})F_t^Z \qquad (1)$$
$$f_t^z = \sin(-\overline{\Psi_{t-i}^H})F_t^X + \cos(-\overline{\Psi_{t-i}^H})F_t^Z$$

The index i in these equations represents the different steps in the input sequence. A total of 7 input features were selected for our model. First, the current two-dimensional velocity of the head in the horizontal plane (V_{t-i}^X, V_{t-i}^Z) was calculated. Height was not used, as there is no evidence for the relevance of this information with regard to the direction of motion on a plane. These input velocities were rotated using the reference yaw angle to convert them to the head-fixed coordinate system.

$$v_{t-i}^x = \cos(-\overline{\Psi_{t-i}^H})V_{t-i}^X - \sin(-\overline{\Psi_{t-i}^H})V_{t-i}^Z$$
$$\qquad (2)$$
$$v_{t-i}^z = \sin(-\overline{\Psi_{t-i}^H})V_{t-i}^X + \cos(-\overline{\Psi_{t-i}^H})V_{t-i}^Z$$

Second, we added the yaw and pitch angle of the head (Ψ_t^H, Θ_t^H) and the direction of eye gaze (Ψ_t^E, Θ_t^E) to the list of the features. Both features might be informative since humans usually orient their head to the target and direct their gaze to the floor along the future locomotor path. Third, we added the body tracker yaw (Ψ_t^B), which was captured by the additional Vive tracker in the subject's backpack. This provided the model with information about body orientation. The reference angles were subtracted from all angles. Before our final scaling, the angles were set between -180 and 180 degrees, with zero indicating the reference direction.

$$\psi_{t-i}^H = \Psi_{t-i}^H - \overline{\Psi_{t-i}^H}$$
$$\theta_{t-i}^H = \Theta_{t-i}^H - \overline{\Theta_{t-i}^H}$$
$$\psi_{t-i}^B = \Psi_{t-i}^B - \overline{\Psi_{t-i}^H} \qquad (3)$$
$$\psi_{t-i}^E = \Psi_{t-i}^E - \overline{\Psi_{t-i}^H}$$
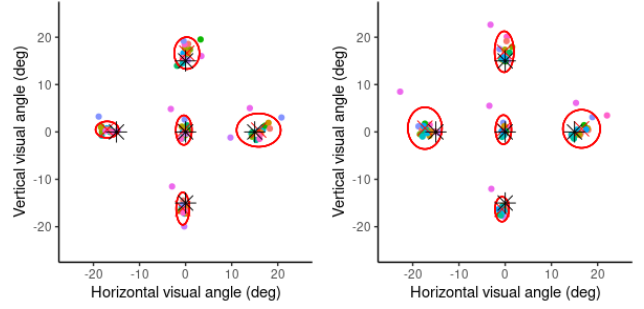$$\theta_{t-i}^E = \Theta_{t-i}^E - \overline{\Theta_{t-i}^H}$$

Figure 2: Comparison of custom eye tracking calibration before (left) and after (right) the experiment. The five black asterisks represent the fixation targets. Red crosses show mean fixation positions across all participants (colored dots) and two standard deviations (ellipses).

The input length was set to 2.5 s (50 samples per input), i.e. the prediction was based on the present data and 2.5 s of past information.

To compensate for possible asymmetries resulting from the architecture of the VE (e.g. the positioning of the room entrance might have led the participants to favor right or left curves in the Search Room), every second data pair was mirrored (left-right).

4.6.2 Processing

The data went through two LSTM layers with 64 hidden units, a dropout layer ($p = 0.3$) and a dense output layer for the predicted coordinates.

We chose adam as the optimizer (learning rate = 0.003) [28]. A weight decay of 1×10^{-4} was added to prevent overfitting. The model was trained for 20 epochs using the mean squared error as the criterion and a batch size of 64.

To obtain a single comprehensible value for estimating the quality of the prediction, we computed the mean displacement error (mde) as the distance between the true value and the prediction. In order to evaluate the influence of eye-tracking and IMU data, we also created models without this data to compare them to the main model, which included all features. The model without IMU data also omitted the eye-tracking features, because a tracking system equipped with an eye tracker but without IMU seemed to be an unlikely use case.

We implemented leave-3-out-cross-validation at the group level. One participant's data were used as validation data and two participants' data were used as test data. All remaining participants were used to train the model. The validation data were used for hyper-parameter optimization. All data were z-standardized with scalers fitted to the training data.

5 RESULTS

Participants took on average 14 minutes to complete the experiment.

5.1 Evaluation of Eye Tracking Stability

The eye tracking calibration analysis revealed an average Euclidean eye tracking error of 1.99 degrees before and 2.07 degrees after the experiment (no significant difference, $p > .05, t(15) = -0.31$) across all tested participants. Thus, the overall error stayed small over the experiment. Within subject tests of fixation positions before and after revealed a significant mean difference of 1.38 degrees ($p < 0.001, t(15) = 5.37$). Thus, individually, the eye tracking changed non-systematically by a small amount, probably due to HMD slippage during natural walking experiment. Centered fixation positions showed standard deviations up to 1.6 degrees. The more peripheral fixation targets at ± 15 degrees had standard deviations up to 2.2 degrees (see Fig. 2).
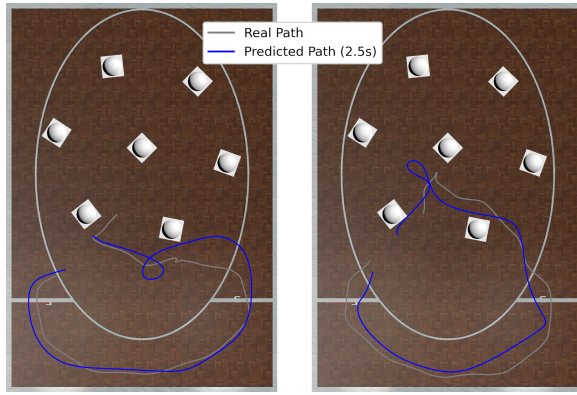
Figure 3: Two examples of 15 seconds of consecutive path predictions. The gray line depicts the real path that the participant walked. The blue line shows the path predicted by the model. Prediction at each data point was based on the preceding 2.5 seconds of movement.

Table 1: 2.5 s prediction error for different rooms

Training Data	Eye Data	Test Data		
		Search	Corridor	Obstacle
All	yes	88 cm	68 cm	59 cm
All	no	90 cm	68 cm	62 cm
Search	yes	84 cm	124 cm	103 cm
Search	no	86 cm	120 cm	101 cm
Corridor	yes	116 cm	61 cm	103 cm
Corridor	no	114 cm	60 cm	97 cm
Obstacle	yes	106 cm	106 cm	57 cm
Obstacle	no	105 cm	111 cm	59 cm
All*	yes	88 cm	75 cm	70 cm

*The number of samples in the Transition Corridor and the Obstacle Room was artificially lowered to match the number of samples in the Search Room.

5.2 Overall Quality of Prediction

The full data set contained 156,076 input-output pairs to train the model. In the 2.5 s labels, the participants travelled an average distance of 165 cm with a mean walking speed of 0.72 m/s. The prediction model using all features (position, orientation and gaze) produced an mde of 66 cm for a 2.5 s prediction. A model using only position and orientation performed slightly worse with an mde = 68 cm. A model that used only position data produced an mde = 78 cm. For comparison, a linear regression model considering all features reached an mde of 93 cm ($\sigma = 8$ cm) and an extrapolation benchmark based on the most recent positions reached an mde of 131 cm ($\sigma = 16$ cm). Figure 3 shows examples of predicted paths from the full model in the Search Room and the Transition Corridor.

Using the testing method proposed by Nadeau and Bengio [37] for cross-validated data (alpha level = 0.05), the difference between the LSTM model and the linear model was significant ($t(5) = -11.08, p < 0.001$). The Benjamini-Hochberg correction [4] was used to test the differences between the three LSTM models. The model using only position data produced significantly larger error than the model using all features ($t(5) = -6.99, p < 0.01$) and the model using only positional and IMU features ($t(5) = -4.92, p < 0.01$). The model using all features produces significantly smaller error than the model using only positional and IMU data ($t(5) = -3.01, p < 0.05$). Thus, eye tracking significantly improved prediction quality, even if only by a small amount (2.78%). Further analysis showed that using a GRU model did not lead to a significantly better prediction (details in [6]).

5.3 Analysis of the Different Rooms

Next, we compared the prediction performances of the model in the three rooms to test whether the improvement gained by eye tracking depends on the task and behavior of the participants.

5.3.1 Search Room

In the Search Room, participants looked for a search object amongst six distractors. Each object had to be closely approached in order to check if it was the search object. On average, the participants walked with a speed of 0.6 m/s and found the search object in the 4th attempt ($\sigma = 0.6$). Typically, after entering the room, participants walked to the object closest to the entrance and approached one of the outer objects (either clockwise or anti-clockwise) next. In most trials, participants then tested the object in the centre, and afterwards continued with the remaining objects (see Figure 4a).

For the model trained with all data, prediction error in the Search Room was higher than the average error across all rooms (see Table 1). One reason could be quantity of training data, since only 14.36 % of the data originated in the Search Room. However, training the model specifically on the Search Room's data improved the prediction only marginally (mde of 84 cm). This suggests that the lower prediction quality in the Search Room might be related to the particular task performed, which might be more difficult to predict than the tasks in the other rooms.

The inclusion of eye data improved the prediction in the Search Room by a small margin, both in the model trained with all the data and in the model specifically trained on the Search Room data. Thus, eye movement data provided a benefit for the prediction in the search task.

5.3.2 Transition Corridor

In the Transition Corridor, participants had to walk along a curve in order to proceed from one room to the other. Participants passed the Transition Corridor at a mean speed of 0.9 m/s.

As the participants went around the curve, they looked towards the inside of the wall and eventually at the door (see Figure 4b). Hence, gaze was to the inside of the curve at almost all times.

The full model trained on all data achieved a prediction error of 68 cm in the Transition Corridor (see Table 1). Training a model specifically on the Transition Corridor data without including data from the other rooms reduces the error to 61 cm, showing a noticeable advantage. 28.56 % of the full model data came from the Transition Corridor.

Eye tracking, on the other hand, did not prove useful in the Transition Corridor. When trained on all data and when trained on only the Transition Corridor data, the model with eye tracking data gave even slightly worse results than the model without eye tracking.

5.3.3 Obstacle Room

In the Obstacle Room, participants had to walk to a target object while avoiding an obstacle along the way (see Figure 1). Participants moved at an average speed of 0.62 m/s through the obstacle room. They passed 90 % of all right obstacles on the right and 87 % of all left obstacles on the left. When the obstacle was placed in the center, 68 % of all obstacles were passed on the right. For the model, this bias should not matter, since half of the data were mirrored.

Every fourth walk contained no obstacle and participants could walk straight to the target. Comparing gaze data in these trials to those trials that contained an obstacle allowed us to look at differences in the eye movements depending on whether the obstacle
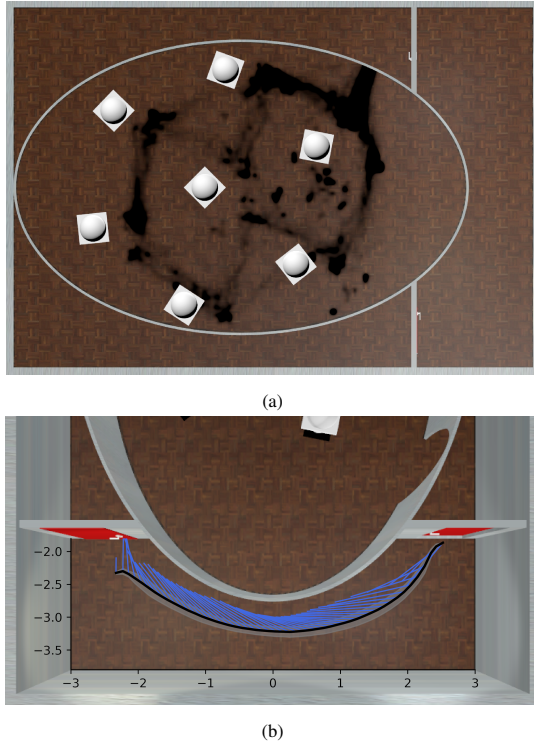
(a)

(b)

Figure 4: Results of the experiments' three parts (all participants). (a) Orthographic Projection of the Search Room with its 7 objects. Each black dot represents a position that one of the participants occupied in that room. The transparency represents path frequency (more frequently walked paths are shown darker). The black spot at the top right shows the entrance to the room. (b) Bird's eye view of the Transition Corridor. The black line is the average walking path of all participants (standard deviation in gray). The blue lines show the mean viewing direction at the respective position.
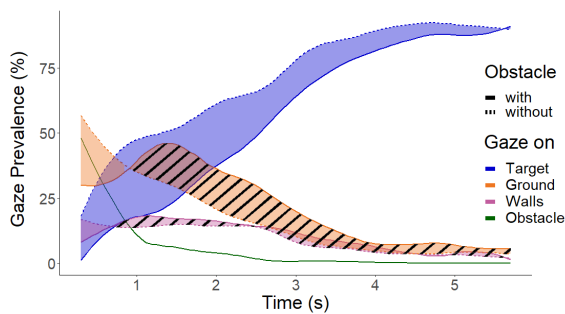

Figure 5: Where did participants look in the Obstacle Room? The colored lines represent the relative prevalence of looking at the target, ground, wall, or obstacle as participants walked from the buzzer to the target. Conditions with and without obstacle are indicated by continuous and dashed lines, respectively. Colored areas between the lines identify phases in which participants looked to the respective object more often when the obstacle was absent than when it was present. Shaded areas indicate phases in which participants looked to the respective object more often when the obstacle was present than when it was absent.

was present or not. Figure 5 shows the proportion of gazes (gaze prevalence) that were directed to the target (blue), obstacle (green), ground (orange), or walls (pink) as a function of time from the start of the walk. Conditions with an obstacle are plotted as lines, conditions without an obstacle as dashed lines. The figure shows that, at the start of the walk, within the first second, participants looked mostly at the ground. Then, when no obstacle was present, they looked at the target object and kept their gaze there most of the time. When an obstacle was present, participants looked less frequently to the target, partially because they looked at the obstacle, but also because they looked more often to the ground than when no obstacle was present. Overall, over the course of walking to the target, our participants looked progressively more often at the target and less often at the ground. Right after start, half of all gazes were directed to the obstacle when an obstacle was present. Gazes to the obstacle became less frequent after three seconds. At that time, in many trials, the obstacle had been passed and was no longer visible.

The mde in the Obstacle Room (59 cm) was smaller than the mean error over all rooms (see Table 1). When the model was trained specifically on Obstacle Room data, the error was even smaller (57 cm) 57.08 % of the data originated in the Obstacle Room.

The inclusion of eye data provided a small advantage to predictions, both when data from all rooms was used for training and when only the Obstacle Room was used.

5.3.4 Comparison Between Rooms

When comparing the rooms, the best prediction result was achieved in the Obstacle Room, followed by the Transition Corridor and the Search Room. While there were differences in the amount of data collected in different rooms, the prediction results still hold when the amount of data were artificially reduced to the same number of observations in all rooms (see Table 1). Training the model on only a single room's data slightly improved the predictions in that room at the expense of prediction accuracy in the other rooms.

5.4 Dependence of Eye-Tracking Benefit on Locomotor and Gaze Behavior

Eye movements provided the biggest benefits in the Search and Obstacle Room. This indicates that task and behaviour can influence the importance of eye tracking data for the model.

Figure 6a shows the improvement of prediction error for a model with eye movements compared to a model without eye movements as a function of the mean acceleration in the 2.5 s input data segment. Colors indicate the different rooms and transparency shows the amount of data that was available. Prediction quality improved when acceleration was larger, most notably in the Obstacle Room. This shows that eye movements were particularly useful when the data segment contained high acceleration. Figure 6b shows how the prediction error varies with the mean acceleration in the 2.5 s data segment that was to be predicted. Here, prediction quality was best when the user decelerated in the Obstacle Room. This suggests that eye movements are most useful for prediction in situations in which the user is likely to stop, presumably because she is close to the target. The Search Room and the Transition Corridor did not show such dependencies. However, in these rooms, large accelerations and decelerations were less likely to occur, because they were not part of the typical task-related behavior in this room.

Besides the acceleration in the walking data, gaze data were also useful to indicate situations in which eye tracking reduces the prediction error. Figure 6c shows the improvement of prediction error for a model with eye movements compared to a model without eye movements as a function of the gaze distance in the 2.5 s input data segment, i.e. the distance to the object at which the user currently looks. In the Obstacle Room, prediction quality benefits the most from eye tracking when the gaze target is close. This tendency was also present in the Search Room but not in the Transition Corridor.
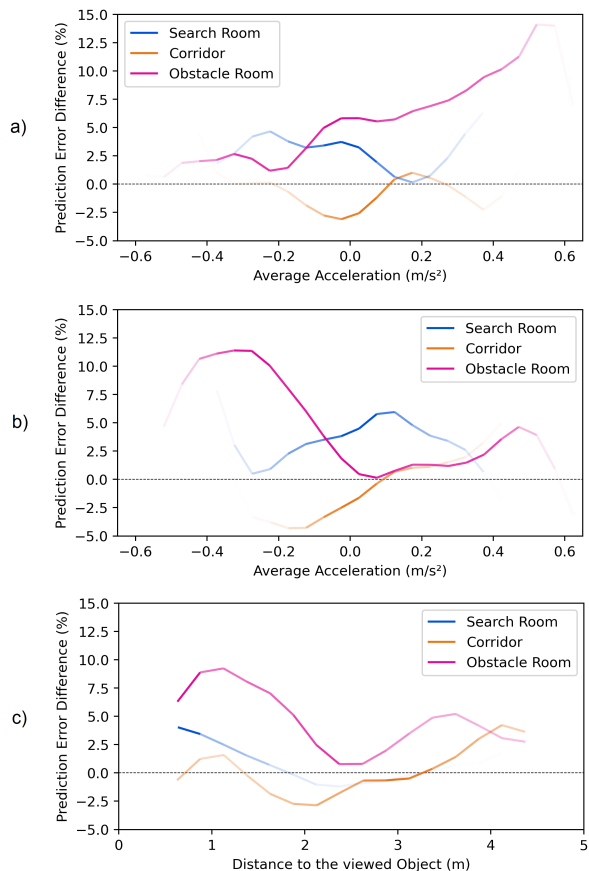
Figure 6: Added value of eye tracking in the model. (a) shows the model advantage as a function of the acceleration in the input sequence position data. A 0.25 m/s² long Gaussian rolling window (standard normal distribution) was used to smooth the data. (b) shows the model advantage as a function of the acceleration in the true output paths. The same rolling window was used. (c) shows the model advantage as a function of the gaze length, i.e. the distance between the observer and the point where their gaze hits the world, at the moment of prediction. A 1.25 m long Gaussian rolling window was used. The transparency of the line colors indicates the number of observations that factored into each data point.

5.5 Questionnaires

The mean SSQ total simulator sickness increased from 7.48 ($\sigma = 7.70$) before the experiment to 20.36 ($\sigma = 34.55$) after the experiment. However, mean values around 20 should not be automatically attributed to a bad simulator in novice VR users [5]. Therefore and because we did not notice a high occurrence of motion sickness, when talking to the participants after the experiment, we would interpret the value increase as a typical result from an experiment that included a 14 minutes long physical task (walking with a VR backpack) and was done by novice users. Moreover, such physical activity has led to increased sweating, which is also represented as a sickness-symptom in the SSQ.

The results from the SUS questionnaire indicated that the users perceived the presented VE as immersive. Participants scored an average of 4.99 ($\sigma = 1.28$, $min = 2.83$, $max = 7$).

6 Discussion

We investigated whether data obtained from eye tracking devices in current VR hardware can be used to enhance an LSTM model of locomotor path prediction for natural walking in VR. The model predicted future walking positions trained on position, orientation, and eye tracking data from a free walking scenario, in which users performed different tasks in differently structured rooms. The full model produced good predictions that exceeded those of linear regression and simple extrapolation models.

To evaluate the use of eye tracking in this model, we compared it against models that used only position and orientation, or only position data for training. The model including eye tracking significantly improved prediction compared to those models, albeit only to a small amount overall. We evaluated the impact of eye tracking data in the model specifically for the different rooms, tasks, and walking behaviours. We found eye tracking benefits in search and obstacle avoidance tasks and especially in situations in which users changed their walking velocity.

6.1 Quality of Eye Tracking in VR

A prerequisite for the predictive utility of eye data in VR is the quality and stability of the eye tracker. Different currently available eye trackers have different temporal resolutions, making them more or less usable for studies needing temporal accuracy in the millisecond range [45]. For the present study, temporal accuracy was less critical since we averaged data over 50 ms intervals and based our prediction on data segments of 2.5 s. Since locomotor behavior in walking is usually smooth, a small temporal lag of the eye tracker would not be detrimental in our prediction scenario. However, during extended periods of walking, on average 14 minutes in our study, spatial accuracy of the eye tracking might deteriorate, for example, if the HMD slips on the head during user movement.

To check eye tracking quality, we measured fixations to a standard set of 5 fixation targets before and after the experiment. We evaluated both the average error before and after the experiment and the within-subject change of the fixation positions before and after. There was an average error of 2 degrees in the eye tracking data within our 30 degree field of view, which did not increase significantly after the experiment. The average within-subject change of the fixation positions was 1.38 degrees. Therefore, we consider the stability of the eye tracking suitable for our purpose of measuring general directions of gaze in walking experiments. While this is encouraging for the use of eye tracking in VR, we believe that applications that require eye positions should include such a before-and-after calibration check.

6.2 Gaze and Locomotor Behavior in VR

Gaze behavior during locomotion in the real world shows certain characteristics, like looking on the ground in front in cluttered environments [9, 36, 41, 49], looking at the target in approach [13, 22], or looking towards the inside of a curve [17, 26]. The observed behavior in the different rooms is consistent with these specifics, suggesting that participants acted naturally in these virtual tasks.

The walking speed in the Transition Corridor was higher than in the other rooms. This might be explained by a lack of task-relevant object interaction, which allowed the participants to pass through quickly. In the Search and Obstacle Rooms, participants walked more slowly and interacted with objects more deliberately. In the Obstacle Room, participants glanced more frequently to the ground in the first three seconds of their way to the target when an obstacle was present, consistent with looks to the ground in natural behavior for negotiating a path containing obstacles [9, 36, 41, 49]. As the obstacle blocked the view towards the target, participants may have had to look sideways towards their path. Since gaze precedes path [e.g. 21, 36], it would also be possible that curves to avoid the obstacle necessitated smaller shifts of the gaze here. In the Search Room, the systematic search patterns shown in Figure 4a revealed that short distances were preferred to arbitrary trial and error.

6.3 Use of Eye Tracking for Prediction

Our results show that eye tracking can help estimating future actions. Gandrud and Interrante [16] previously showed that gaze data are useful when predicting choice between two potential walking targets in a VR hallway. Likewise, Zank and Kunz [56] identified eye data as especially advantageous for binary walking path predictions in long narrow environments. We expanded their results by following a more general approach, not limited to binary predictions. This provide the potential to be applicable to a broader range of environments and applications. Users were instructed to walk freely in the VE. Our aim was to predict their future position in space. Therefore, a direct comparison of the performance of our model to the prediction methods in those previous papers is not possible with the current data-set. However, it is possible to compare whether eye tracking data were a significantly useful addition for the prediction in the different studies. Eye movements provide useful information when approaching a target [13, 22]. This fits well with the results of the binary prediction studies. However, when walking in a curve with no task, the head orientation is likely to include the same information as the gaze [17, 26]. Accordingly, gaze provided no benefit for the model using only data from the Transition Corridor.

In the Obstacle Room, an advantage from the inclusion of eye data was observed for decelerating paths and paths that were preceded by accelerated motion. It seems possible that the presence of the obstacle influenced the motion and gaze patterns in a way that eye data improved model performance. However, since the obstacle was only looked at for a short time (see Figure 5), it is not clear if this alone has caused this effect. The advantage could also be caused by a change in gaze behavior when stopping in front of the target. However, the Obstacle Room was the only room where stopping was part of the task. Therefore, we cannot clearly differentiate if eye movements were most beneficial for predictions during stop-and-go, obstacle avoidance or a combination of both tasks.

The link between gaze and locomotor activity in natural behavior and VR suggests that eye tracking might be useful for predicting upcoming user actions. The results from our LSTM prediction model confirm that the addition of eye tracking data can provide a significant benefit in predicting future walking paths. Overall, the mde of the best model including gaze data was 66 cm for a 2.5 s prediction. The average distance walked during 2.5 s was 165 cm. It remains to be evaluated in future studies how effective this prediction is for different applications.

6.4 Possible Application Scenarios

For walking in VR, a particular application scenario is RDW, in which users are steered along a physical trajectory that differs from the virtual trajectory in order to optimally use the physical space available for tracking [42]. In dynamic RDW, one has to decide in which direction and by how much the user should be redirected at any given point in time. Knowledge of where the user most likely intends to go can be advantageous for implementing RDW fast and effectively [40]. Since our prediction model worked well in different rooms and tasks and since it only uses data directly from the user, it can be considered a useful tool for a generalized RDW controller. Future user studies are needed to evaluate if an RDW controller using eye tracking data for redirection outperforms previous approaches regarding the needed number of resets and the user experience in different physical and VR scenarios.

A very simple way to include our prediction in RDW is to focus on whether a user, at a particular point in time, should be redirected leftward or rightward. In such a case, a simple left/right prediction of the users intent could be sufficient. To quantify the quality of such a prediction in our model, we took the predicted future positions at each point in time and subjected them to a simple left/right discrimination. The results showed that the model could predict whether the user will turn left or turn right with an accuracy of 84.8 %. A model

explicitly trained to distinguish between left curves and right curves would likely achieve even better accuracy. Additionally, there might be some potential for synergistic effects with other methods, such as following behavior [39] to improve performance even further.

Our method of locomotion prediction might also be useful in other scenarios. For example, control of non-player characters could benefit from user prediction to avoid collisions. Moreover, valid predictions of upcoming behavior could be helpful in determining the objects and locations where computational resources should be focused to e.g. increase their responsiveness or level of detail.

An important feature of our method is that it relies exclusively on egocentric data and does not need information about the environment. This feature makes the method very versatile. By using inside-out tracking, head worn IMUs and eye trackers, the data for prediction can be gathered by sensors worn by the user, not only in VR. For example, similar models could likely be trained to provide rough predictions of locomotion intentions in augmented reality scenarios. Moreover, the anticipation of human actions, such as walking, can also play a key role in the development of assistive robots [30].

6.5 Limitations

Although we tried to create an experiment with a representative sample of typical VR walking tasks, our data set is limited to our selection of rooms and tasks. This could compromise the performance of our model in new contexts, as movements associated with our tasks could be disproportionately represented in the data set. This effect is already observable in our results, since the models that limited access to some rooms during training produced worse results in the other rooms, compared to the full model (see Table 1). While the overall smaller amount of data might have decreased prediction accuracy, the discrepancy between known and unknown environments indicates that the transfer from one of our rooms to the others has only been successful to a limited extent. This suggests that performance of our model is likely to drop when used with tasks that were not included in the training phase. However, even these higher errors are still in a range far below the distances traveled. Therefore, we think that our model's path prediction can be used in new situations as long as an error of about 1 m in comparison to the real path is tolerable for the application.

Another limitation is concerned with the predicted variable we used in our model. Minimizing the error between the real and the predicted end position of a 2.5 s subset of data gave us valuable information about the range of mde and the incremental value of gaze data for the prediction. However, it does not include information on the deviation between the predicted and the true trajectory during the walk. Prediction of the full trajectory could be valuable for some application scenarios and therefore should be considered in extended path prediction models in the future.

7 CONCLUSION

Eye tracking data from current VR hardware improves deep learning path prediction for natural walking in VR. Eye tracking benefits appear especially in situations in which the user interacts with the virtual environment during locomotion. Including eye-tracking data while considering the user's task and behavior is a useful tool for deep learning path prediction.

8 DECLARATION OF CONFLICTING INTERESTS

The author(s) declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

9 FUNDING

REFERENCES

[1] G. Arechavaleta, J.-P. Laumond, H. Hicheur, and A. Berthoz. An optimality principle governing human walking. *IEEE Transactions on Robotics*, 24(1):5–14, 2008. doi: 10.1109/TRO.2008.915449

[2] S. Becker, R. Hug, W. Hübner, and M. Arens. An evaluation of trajectory prediction approaches and notes on the trajnet benchmark. *arXiv preprint arXiv:1805.07663*, 2018.

[3] A. Belardinelli, M. Y. Stepper, and M. V. Butz. It's in the eyes: Planning precise manual actions before execution. *Journal of Vision*, 16(1):18–18, 01 2016. doi: 10.1167/16.1.18

[4] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995. doi: 10.1111/j.2517-6161.1995.tb02031.x

[5] P. Bimberg, T. Weissker, and A. Kulik. On the usage of the simulator sickness questionnaire for virtual reality research. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 464–467, 2020. doi: 10.1109/VRW50115.2020.00098

[6] G. Bremer, N. Stein, and M. Lappe. Predicting future position from natural walking and eye movements with machine learning. In *2021 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pp. 19–28. IEEE, 2021. doi: 10.1109/AIVR52153.2021.00013

[7] H. Brument, I. Podkosova, H. Kaufmann, A. H. Olivier, and F. Argelaguet. Virtual vs. physical navigation in vr: Study of gaze and body segments temporal reorientation behaviour. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 680–689. IEEE, 2019. doi: 10.1109/VR.2019.8797721

[8] L. Bölling, N. Stein, F. Steinicke, and M. Lappe. Shrinking circles: Adaptation to increased curvature gain in redirected walking. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):2032–2039, 2019. doi: 10.1109/TVCG.2019.2899228

[9] D. Calow and M. Lappe. Efficient encoding of natural optic flow. *Network Comput. Neural Syst.*, 19(3):183–212, 2008. doi: 10.1080/09548980802368764

[10] Y.-H. Cho, D.-Y. Lee, and I.-K. Lee. Path prediction using lstm network for redirected walking. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 527–528. IEEE, 2018. doi: 10.1109/VR.2018.8446442

[11] V. Clay, P. König, and S. Koenig. Eye tracking in virtual reality. *Journal of Eye Movement Research*, 12(1), 2019. doi: 10.16910/jemr.12.1.3

[12] M. Cornia, L. Baraldi, G. Serra, and R. Cucchiara. Predicting human eye fixations via an lstm-based saliency attentive model. *IEEE Transactions on Image Processing*, 27(10):5142–5154, 2018. doi: 10.1109/TIP.2018.2851672

[13] S. Durant and J. M. Zanker. The combined effect of eye movements improve head centred local motion information during walking. *PloS one*, 15(1):e0228345, 2020. doi: 10.1371/journal.pone.0228345

[14] X. Feng, Y. Liu, and S. Wei. Livedeep: Online viewport prediction for live virtual reality streaming using lifelong deep learning. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 800–808, 2020. doi: 10.1109/VR46266.2020.00104

[15] P. W. Fink, P. S. Foo, and W. H. Warren. Obstacle avoidance during walking in real and virtual environments. *ACM Transactions on Applied Perception (TAP)*, 4(1):2–es, 2007. doi: 10.1145/1227134.1227136

[16] J. Gandrud and V. Interrante. Predicting destination using head orientation and gaze direction during locomotion in vr. In *ACM Symposium on Applied Perception, SAP 2016*, pp. 31–38. Association for Computing Machinery, Inc, 2016. doi: 10.1145/2931002.2931010

[17] R. Grasso, P. Prévost, Y. P. Ivanenko, and A. Berthoz. Eye-head coordination for the steering of locomotion in humans: an anticipatory synergy. *Neuroscience Letters*, 253(2):115–118, 1998. doi: 10.1016/S0304-3940(98)00625-9

[18] M. Hayhoe and D. Ballard. Eye movements in natural behavior. *Trends Cogn. Sci.*, 9(4):188–194, 2005. doi: 10.1016/j.tics.2005.02.009

[19] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735

[20] M. A. Hollands and D. E. Marple-Horvat. Visually guided stepping under conditions of step cycle-related denial of visual information. *Experimental brain research*, 109(2):343–356, 1996. doi: 10.1007/BF00231792

[21] M. A. Hollands, D. E. Marple-Horvat, S. Henkes, and A. K. Rowan. Human eye movements during visually guided stepping. *Journal of motor behavior*, 27(2):155–163, 1995. doi: 10.1080/00222895.1995.9941707

[22] M. A. Hollands, A. E. Patla, and J. N. Vickers. "look where you're going!": gaze behaviour associated with maintaining and changing the direction of locomotion. *Experimental brain research*, 143(2):221–230, 2002. doi: 10.1007/s00221-001-0983-7

[23] Z. Hu, A. Bulling, S. Li, and G. Wang. Fixationnet: Forecasting eye fixations in task-oriented virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2681–2690, 2021. doi: 10.1109/TVCG.2021.3067779

[24] Z. Hu, S. Li, C. Zhang, K. Yi, G. Wang, and D. Manocha. Dgaze: Cnn-based gaze prediction in dynamic scenes. *IEEE Transactions on Visualization and Computer Graphics*, 26(5):1902–1911, 2020. doi: 10.1109/TVCG.2020.2973473

[25] C. Hutton and E. Suma. A realistic walking model for enhancing redirection in virtual reality. In *2016 IEEE Virtual Reality (VR)*, pp. 183–184. IEEE, 2016. doi: 10.1109/VR.2016.7504714

[26] T. Imai, S. T. Moore, T. Raphan, and B. Cohen. Interaction of the body, head, and eyes during walking and turning. *Experimental brain research*, 136(1):1–18, 2001. doi: 10.1007/s002210000533

[27] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993. doi: 10.1207/s15327108ijap0303_3

[28] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[29] D. Kit, L. Katz, B. Sullivan, K. Snyder, D. Ballard, and M. Hayhoe. Eye movements, visual search and scene memory, in an immersive virtual environment. *PLOS ONE*, 9(4):1–11, 04 2014. doi: 10.1371/journal.pone.0094362

[30] H. S. Koppula and A. Saxena. Anticipating human activities using object affordances for reactive robotic response. *IEEE transactions on pattern analysis and machine intelligence*, 38(1):14–29, 2015. doi: 10.1109/TPAMI.2015.2430335

[31] M. Land and B. Tatler. *Locomotion on foot*, pp. 100–115. Oxford University Press, 07 2009. doi: 10.1093/acprof:oso/9780198570943.003.0006

[32] M. F. Land and M. Hayhoe. In what ways do eye movements contribute to everyday activities? *Vision research*, 41(25-26):3559–3565, 2001. doi: 10.1016/S0042-6989(01)00102-X

[33] E. Langbehn, P. Lubos, and F. Steinicke. Evaluation of locomotion techniques for room-scale vr: Joystick, teleportation, and redirected walking. In *Proceedings of the Virtual Reality International Conference-Laval Virtual*, pp. 1–9, 2018. doi: 10.1145/3234253.3234291

[34] D. J. Lohr, L. Friedman, and O. V. Komogortsev. Evaluating the data quality of eye tracking signals from a virtual reality system: Case study using smi's eye-tracking htc vive. *arXiv preprint arXiv:1912.02083*, 2019.

[35] D. S. Marigold and A. E. Patla. Visual information from the lower visual field is important for walking across multi-surface terrain. *Experimental Brain Research*, 188(1):23–31, 2008. doi: 10.1007/s00221-008-1335-7

[36] J. S. Matthis, J. L. Yates, and M. M. Hayhoe. Gaze and the control of foot placement when walking in natural terrain. *Current Biology*, 28(8):1224–1233, 2018. doi: 10.1016/j.cub.2018.03.008

[37] C. Nadeau and Y. Bengio. Inference for the generalization error. *Machine learning*, 52(3):239–281, 2003. doi: 10.1023/A:1024068626366

[38] T. Nescher, Y.-Y. Huang, and A. Kunz. Planning redirection techniques for optimal free walking experience using model predictive control. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 111–118. IEEE, 2014. doi: 10.1109/3DUI.2014.6798851

[39] A. Nguyen, P. Wüest, and A. Kunz. Human following behavior in virtual reality. In *26th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–3, 2020. doi: 10.1145/3385956.3422099

[40] N. C. Nilsson, T. Peck, G. Bruder, E. Hodgson, S. Serafin, M. Whitton, F. Steinicke, and E. S. Rosenberg. 15 years of research on redirected walking in immersive virtual environments. *IEEE computer graphics and applications*, 38(2):44–56, 2018. doi: 10.1109/MCG.2018.111125628

[41] A. E. Patla and J. N. Vickers. Where and when do we look as we approach and step over an obstacle in the travel path? *Neuroreport*, 8(17):3661–3665, 1997. doi: 10.1097/00001756-199712010-00002

[42] S. Razzaque, Z. Kohn, and M. C. Whitton. Redirected Walking. In *Eurographics 2001 - Short Presentations*. Eurographics Association, 2001. doi: 10.2312/egs.20011036

[43] A. Sipatchin, S. Wahl, and K. Rifai. Eye-tracking for clinical ophthalmology with virtual reality (vr): A case study of the htc vive pro eye's usability. In *Healthcare*, vol. 9, p. 180. Multidisciplinary Digital Publishing Institute, 2021. doi: 10.3390/healthcare9020180

[44] M. Slater, M. Usoh, and A. Steed. Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 3(2):130–144, 1994. doi: 10.1162/pres.1994.3.2.130

[45] N. Stein, D. C. Niehorster, T. Watson, F. Steinicke, K. Rifai, S. Wahl, and M. Lappe. A comparison of eye tracking latencies among several commercial head-mounted displays. *i-Perception*, 12(1):2041669520983338, 2021. doi: 10.1177/2041669520983338

[46] F. Steinicke, G. Bruder, J. Jerald, H. Frenz, and M. Lappe. Estimation of detection thresholds for redirected walking techniques. *IEEE transactions on visualization and computer graphics*, 16(1):17–27, 2009. doi: 10.1109/TVCG.2009.62

[47] F. Steinicke, Y. Visell, J. Campos, and A. Lécuyer. *Human walking in virtual environments*, vol. 2. Springer, 2013. doi: 10.1007/978-1-4419-8432-6

[48] E. A. Suma, Z. Lipps, S. Finkelstein, D. M. Krum, and M. Bolas. Impossible spaces: Maximizing natural walking in virtual environments with self-overlapping architecture. *IEEE Transactions on Visualization and Computer Graphics*, 18(4):555–564, 2012. doi: 10.1109/TVCG.2012.47

[49] B. M. 't Hart and W. Einhauser. Mind the step: complementary effects of an implicit task on eye and head movements in real-life gaze allocation. *Exp. Brain Res.*, 223(2):233–249, 2012. doi: 10.1007/s00221-012-3254-x

[50] B. W. Tatler and S. L. Tatler. The influence of instructions on object memory in a real-world setting. *Journal of Vision*, 13(2):5–5, 02 2013. doi: 10.1167/13.2.5

[51] S. Tuhkanen, J. Pekkanen, P. Rinkkala, C. Mole, R. M. Wilkie, and O. Lappi. Humans use predictive gaze strategies to target waypoints for steering. *Scientific reports*, 9(1):1–18, 2019. doi: 10.1038/s41598-019-44723-0

[52] M. Usoh, K. Arthur, M. C. Whitton, R. Bastos, A. Steed, M. Slater, and F. P. Brooks Jr. Walking >walking-in-place >flying, in virtual environments. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pp. 359–364, 1999. doi: doi.org/10.1145/311535.311589

[53] J. Wiener, O. De Condappa, and C. Holscher. Do you have to look where you go? gaze behaviour during spatial decision making. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 33, 2011.

[54] Y. Xu, Y. Dong, J. Wu, Z. Sun, Z. Shi, J. Yu, and S. Gao. Gaze prediction in dynamic 360° immersive videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. doi: 10.1109/CVPR.2018.00559

[55] C. Yu, X. Ma, J. Ren, H. Zhao, and S. Yi. Spatio-temporal graph transformer networks for pedestrian trajectory prediction. In *European Conference on Computer Vision*, pp. 507–523. Springer, 2020. doi: 10.1007/978-3-030-58610-2_30

[56] M. Zank and A. Kunz. Eye tracking for locomotion prediction in redirected walking. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 49–58. IEEE, 2016. doi: 10.1109/3DUI .2016.7460030

[57] M. Zank and A. Kunz. Where are you going? using human locomotion models for target estimation. *The Visual Computer*, 32(10):1323–1335, 2016. doi: 10.1007/s00371-016-1229-9

[58] M. Zank and A. Kunz. Optimized graph extraction and locomotion prediction for redirected walking. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 120–129. IEEE, 2017. doi: 10.1109/3DUI.2017.7893328

[59] M. A. Zmuda, J. L. Wonser, E. R. Bachmann, and E. Hodgson. Optimizing constrained-environment redirected walking instructions using search techniques. *IEEE transactions on visualization and computer graphics*, 19(11):1872–1884, 2013. doi: 10.1109/TVCG.2013.88