Deep-learning based recommenders for the improved user navigation in VR

Murtada Dohan, Mu Mu, Suraj Ajit
The University of Northampton
Waterside Campus
Northampton,UK

Tawfiq A. Al-Assadi
University of Babylon
Babylon, Iraq

Gary Hill
Cranfield University
Cranfield, UK

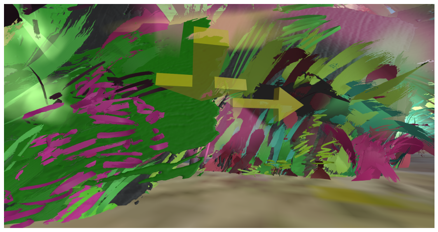
Andreas Mauthe
University of Koblenz
Koblenz, Germany

Abstract

Virtual reality (VR) has become a popular choice for education, industrial simulation, entertainment and healthcare applications. User navigation is an essential propriety of virtual applications. However, novice audiences often face the difficulty of engaging with the virtual surrounding environment. This work presents a novel design of a deep learning-based navigation solution to improve the quality of user experience and the engagement with virtual content. We compare two navigation methods avatar-based and arrows-based guidance, both of which are driven by a recurrent neural network (RNN) model. We capture participants' mobility and eye-gaze to compare the impact of different navigation affects on users' engagement in VR applications.

1 Introduction

Developers and researchers are encouraged to adopt a wide range of virtual reality (VR) tools and applications that help in creating more immersive content. The availability of economical VR headsets comes with a variety of data



(a) Arrows Scene

(b) Avatar Scene

Figure 1: Arrow and avatar-based navigation

providers which help developers and researchers to create more attractive content for the community. However, many users are unable to navigate and explore their surrounding environments due to the challenges created by VR applications [Alghofaili et al.(2019)]. This affects the engagement of the participants with the VR applications, especially for people with limited experiences in VR.

In this research, we introduce a machine learning-driven navigation model that recommends directions in VR based on patterns of users' movements. This model combines different data analysis and machine learning techniques to provide guidance autonomously. By observing and learning how VR users interact with the VR environment, the model is able to make recommendations to inexperienced users who are uncertain what to do in VR. The predictive model has been employed in VR scenes using two different methods, namely: directed arrows (an overhead semi-transparent arrow that explicitly recommends users where to go) and live avatars (animated virtual characters making implicit recommendations by walking towards an area of interest). The arrows and avatars are directed based on the machine learning predictions. For the direct arrows, there are two arrows in the scene showing the Top 2 recommendations. While with avatars, three avatars are enabled using the top 3 recommendations from the machine learning model. Each of the avatars/arrows will be directed to a specific destination during playtime, and these elements will keep updating based on participants' mobility. We have set up following hypotheses to be tested in this research which are:

- H1: Machine learning-based recommender can recognise mobility patterns and will help participants to navigate better in the virtual reality.
- H2: Live-avatars (implicit recommendation) are more effective than directed arrows (explicit recommendation) due to the novelty. However, users might be distracted by the activities of avatars.

2 Data Gathering and Modelling

2.1 Experiment

A number of experiments have been conducted to collect data for modelling and testing the proposed model. These experiments have different goals as they are designed for different purposes. The first experiment involved collecting and analysing data for model building. The experiment involved participants freely navigating through an VR artwork by walking in a physical space. There was no explicit goal to be performed until the users explored and comprehended the environment. The VR content used for the experiment was an abstract VR artwork placed on a half side of the hall. Participants start in front of the artwork and are encouraged to explore and navigate the environment. The artwork is created using Google Tilt Brush [Google(2022)], and it consists of thousands of different brushstrokes as shown in Figure 1. The data collected from this experiment were participants' movements, eye-gaze, and hand motions. HTC VIVE PRO EYE [viv(2019)] is used as the main headset for this experiment. This research focuses on mobility modelling; accordingly, the walking data (mobility) are considered for this purpose. The walking data consist of 2D coordinates used to generate the walking path parameters (head_x, head_z), which represent the head position within the experimental surface.

Raw data for each participant were collected separately per frame. Therefore, a high volume of data is gathered during playtime. In this experiment, 35 participants were asked to share feedback. These participants include 20 females and 15 males, and most are aged between 16 and 25. The majority of the participants stated that they do not or only occasionally play computer games. Participants have navigated differently in the virtual environment. Some took shorter paths while other participants preferred to dive more into the artwork and create longer footpath [Murtada Dohan(2022)]. It is worth mentioning that the brushstrokes of the artwork are non-collided objects so that participants can walk through them to explore the inside of the artwork.

To prepare data for modelling, data processing is an essential operation. The dataset was gathered from multiple sensors at various time points during the experiment. One of the priorities is to ensure that the data from all devices are synchronised. The synchronisation level ensures the timestamp is fixated over multi-sensor capture. We translated all the data into game time to correlate participants' actions with their interactivity. Several pre-processing steps are applied to the data to prepare it for the following processing stages: handling missing values and removing anomalies, biases, and outliers. It is noteworthy that Human-data processing is a complex task. The collection of such data usually takes place in an experimental environment. For example, in the context of data cleaning, redundant or repetitive data were removed, and the data format is changed as needed. As an alternative to extrapolating missing values using filling techniques, we considered removing incomplete data from the dataset as filling techniques are inappropriate for human activity data.

Following the data processing stage, the walking data have been modelled into sequential paths that match the actual paths, which users originate during their navigation. The paths are constructed in directed lines, which show the user's direction among the areas he\she explored. We have divided the virtual environment into many sub-areas using K-means clustering. The clusters maintain some criteria for the size of the cluster and the amount of data in each cluster. The clusters also indicate the most visited areas in the VR environment.

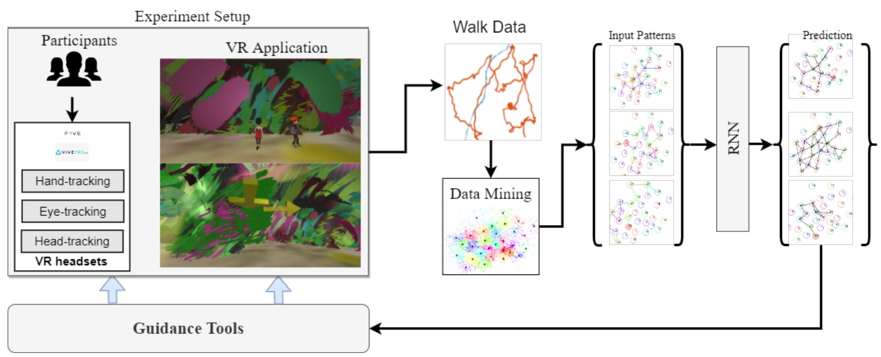


Figure 2: Navigation's Model, model considers participants movements as input and feedback the output to the VE

The model groups the data based on 30 selected areas in the virtual environment. These are accessible to the users by free walking. The number of clusters is chosen according to some tests that have been done in relation to the previous research. Each participant walking into a clustering area in the scene will belong to one of the already assigned clustering areas. Following the clustering, a deep learning model considers the sequences of clusters that participants visited as an input in order to predict the next destination (which represents a cluster). The data neutrality in this model is timeseries data since the sequence of clusters that generates a path is defined as a group of timesteps. The model structure was built using a recurrent neural network (RNN) employing Long Short Term Memory (LSTM) layers [Murtada Dohan(2022)], as depicted in Figure 2

3 Navigation

The Navigation recommender was implemented based on a pre-built AI model that was briefly explained in the previous section 2.1. The navigation model collects walking data in real-time to predict the best destination in the virtual environment. The walking data are represented by the floor coordinates; then, the data were preprocessed and formulated to fit the proposed model. The model considers participants' walk history to predict the upcoming destinations that participants are willing to visit. The model was evaluated using unseen data to test the prediction's accuracy.

The recommender was implemented using two different techniques. The two scenes for implementing the navigation used the same environmental setup except for the guidance elements as shown in Figure 1. Scene-One uses directed arrows, in which two semi-transparent arrows were placed above the head of the participants. The arrows change direction in real-time based on users' location and mobility. In comparison, the other scene uses avatars, which walk and explore the VR scene based on the AI model. We have used three avatars that navigate the environment using Top 3 recommendations from the trained model. An experiment was conducted to evaluate the guidance tools and recommender performance for these purposes. The same virtual environment that is used for data collection and modelling in Section 2.1 is used for this experiment. Additionally, guidance tools are added to the scene to integrate with the pre-trained AI model to recommend walk directions in the virtual environment. 15 participants provided feedbacks to the model's recommendations during the evaluation experiment. Participants were composed of 3 females and 12 males. The participants had similar experiences as in the previous experiment. Participants started the experiment at a random scene (Arrows or Avatars) as shown in Figures 1a and 1b.

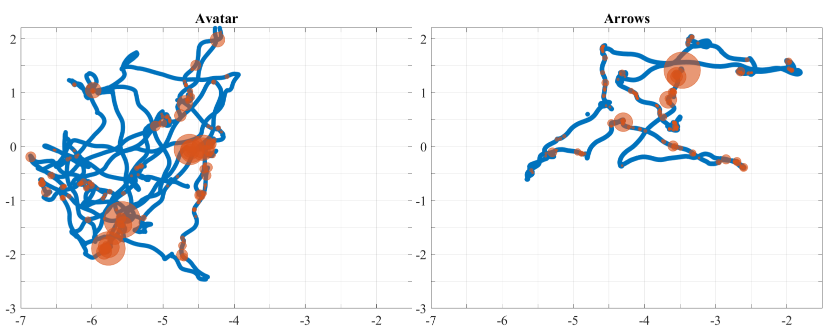


Figure 3: Gaze disruption in Avatars and Arrows scenarios

The figure 3 belong to the same participant who has taken part in both of the evaluation scenes. The blue line refers to the user walking path in virtual space, while the red spot refers to the most areas with eye gaze. It is clear that the participant has explored more areas with the avatar scene than in the Arrows scene, as the path with avatars is longer than the path in the arrow scene. More evaluations are being carried out to further verify our hypotheses.

4 Conclusion

The current paper shows provisional results of implementing a navigation recommender model based on actual walk data. The model is embedded in the same VR environment used to collect the data as we are looking to investigate the influence of navigation guidance on participants. Also, we have employed two different techniques to illustrate these recommendations to the participants via directed arrows and Avatars. Visual observations show that people shared the scene with moving avatars would have engaged more with the content by navigating and using eye-gaze analysis as shown in Figure 3, which opposes the H2. It is clear to conclude Arrows and Avatars have different impacts on most participants who joined the evaluation trial. Out of 15 participants, 12 have been heavily engaged in the Avatar scene rather than the others, which meets H1. The future directions of this research are to statistically assess the influence of recommenders on participants' navigation.

References

[viv(2019)] 2019. HTC VIVE Pro Eye. https://www.vive.com/uk/product/vive-pro-eye/

[Alghofaili et al.(2019)] 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 *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[Google(2022)] Google. 2022. Tilt Brush. *Tilt Brush by Google* (2022). https://www.tiltbrush.com/

[Murtada Dohan(2022)] Gary Hill Murtada Dohan, Mu Mu. 2022. Real-Walk Modelling: Deep Learning Model for User Mobility in Virtual Reality. *Submitted at Transactions on Multimedia Computing Communications and Applications* (2022). submitted.