REGULAR PAPER

Real-walk modelling: deep learning model for user mobility in virtual reality

Murtada Dohan[1] · Mu Mu[1] · Suraj Ajit[1] · Gary Hill[2]

Abstract

This paper presents a study on modelling user free walk mobility in virtual reality (VR) art exhibition. The main objective is to investigate and model users' mobility sequences during interactions with artwork in VR. We employ a range of machine learning (ML) techniques to define scenes of interest in VR, capturing user mobility patterns. Our approach utilises a long short-term memory (LSTM) model to effectively model and predict users' future movements in VR environments, particularly in scenarios where clear walking paths and directions are not provided to participants. The DL model demonstrates high accuracy in predicting user movements, enabling a better understanding of audience interactions with the artwork. It opens avenues for developing new VR applications, such as community-based navigation, virtual art guides, and enhanced virtual audience engagement. The results highlight the potential for improved user engagement and effective navigation within virtual environments.

Keywords Virtual reality · Spatial knowledge · Movement · Deep learning · Navigation · Dataset

1 Introduction

VR tends to act the reality; most applications seek to simulate the real world by adding more elements and visualisation. The development of VR headsets has extended the virtual environment (VE) that embeds more human characteristics inside VR, such as eye tracking, hand gestures, and human movements. The unrestricted development of VR headsets towards reality adds more physical characteristics to VE [1].

Within a typical virtual reality environment, users can employ various navigation methods to move around. These methods include teleportation and controller-based movement, as well as real walking. Real walking (RW) involves physically walking in the real world while wearing a VR headset. The user's movements are tracked and translated into movements within the virtual environment, allowing for a natural and intuitive way to navigate that which closely mimics the biomechanics of RW. Therefore, navigation is a fundamental attribute of user interaction in VR. Research shows that the effectiveness of VR navigation is often determined by the user's previous experience [2]. A common challenge of VR navigation is users getting lost while exploring in an open VE. Machine learning (ML) approaches have been used to improve users' navigating experience in VR. For example, Alghofaili et al. [3] developed a DL model to predict when users need navigation help and adaptively aid them in finding the right way. The results demonstrated the potential in improving the engagement of users in virtual navigation while effectively guiding them to their destinations.

The RW is a vital aspect of the immersive experience which adds more reality sense to the VE [4, 5]. It is an

✉ Murtada Dohan
  murtada.dohan@northampton.ac.uk

✉ Mu Mu
  mu.mu@northampton.ac.uk

  Suraj Ajit
  Suraj.Ajit@northampton.ac.uk

  Gary Hill
  DrGaryHill@outlook.com

[1] Faculty of Arts, Science & Technology, University of Northampton, Waterside Campus, Northampton NN1 5PH, Northamptonshire, UK

[2] MKU, Cranfield University, College Road, Cranfield, Bedfordshire, UK

immersive mobility approach for navigating VR, enhancing the sense of presence for participants [2, 6, 7]. Therefore, modelling users' real work is a crucial aspect for studying user interactions in VR and supporting the future development of intelligent applications that can adapt to users' needs and preferences.

We piloted a VR real walk study in the context of fine art VR painting exhibitions where the audience can explore paintings using RW. To this end, we teamed up with a VR artist and developed a mobility experiment based on a large-scale abstract VR painting. The study's main goal is to investigate and model human movements while RW interactions with a VR artwork take place. Therefore, we involved a range of human-related features in this experiment. The VR painting was created by artist Goodyear using Google Tilt Brush [8]. The experimental environment was developed using a Unity3D game engine with a combination of hardware sensors and software tracking tools. It enables us to collect user data using eye gaze and body movement tracking capabilities to capture fine-grained user interactions. The experiment was carried out with a group of participants invited to explore an abstract VR painting while freely walking within 4x4 ms of physical space. Collected data include time-coded eye gaze, head orientation, hand movements, mobility, and voice comments.

Our study on user mobility in VR is designed to support the development of human-driven navigation tools for individuals who are new to virtual environments. As a result, these users will be able to navigate the virtual environment more effectively. We collected eye gaze, head elevation, and mobility movements in a purposely designed user experiment. These data have been used to train and evaluate a DL classifier model. The classifier model takes a series of user' steps on the floor space to predict their next step. We discuss the model's prediction accuracy and propose different applications of the model as human-driven navigation for the improved user navigation experience in future VR applications.

The main contributions of this paper include:

- Analyse user movements in an abstract VR painting user experiments.
- Utilisation of ML techniques to define scenes of interest in VR, enabling the modelling of user movement patterns.
- Development and implementation of a DL model to effectively capture and predict participants' mobility during their VR art encounters.

The remainder of this paper is organised as follows. Section 2 discusses the background and related work in VR art, behavioural tracking, and modelling in VR. Section 3 introduces the authors' VR artwork, experimentation system, and user experiment. Data analysis and modelling are discussed in Sects. 4.1 and 5. Section 6 concludes the paper.

2 Background and related work

There is an increasing adoption of alternate reality platforms by content creators and visual artists worldwide [9, 10]. Blortasia is an abstract art world in the sky where viewers fly freely through a surreal maze of evolving sculptures [11]. The authors believe the exploration through art and nature reduces stress, anxiety and inflammation, and has positive effects on attitude, behaviour, and well-being. Hayes, et al. [12] created a virtual replication of an actual art museum with features such as gaze-based main menu interaction, hotspot interaction, and zooming/movement in a 360-degree space. The authors suggested that allowing viewers to look around as they please and focus their attention on the interaction happening between the artwork and the room is something that cannot be easily replicated. In [13], Battisti, et al. presented a framework for a virtual museum based on the use of HTC VIVE. The system allows for movement in the virtual space via controllers as well as walking. A subjective experiment showed that VR, when used in a cultural heritage scenario, requires that the system should be designed and implemented by relying on multi-disciplinary competencies such as arts and computer science.

Pfeuffer et al. [14] investigated body motion as behavioural biometrics for VR to identify a user in the context of authentication or to adapt the VR environment to users' preferences. The authors carried out a user study where participants perform controlled VR tasks including pointing, grabbing, walking, and typing while the system monitored their head, hand, and eye movement data. Classification methods were used to associate behaviour data with users. Furthermore, avatars are commonly used to represent attendees in social VR applications. Body tracking has been used to animate the motions of the avatar based on the body movements of human controllers [15]. In [16], full visuomotor synchrony is achieved using wearable trackers to study implicit gender bias and embodiment in VR. The gender-based eye movement differences in indoor picture viewing was studied using ML classification in [17]. The authors discovered that females have a more extensive search whereas males have more local viewing.

The RW stands as the familiar method of travel for humans. It helps humans to have more sense of the present and naturally navigate the surrounding environment [2, 4, 18]. VR environments help users conceptualise a spatial reality. Different locomotion techniques within the virtual model can influence how people conceptualise a spatial reality [19]. In a VR environment, the effect of locomotion on spatial cognition has already been observed through many

studies. Different navigation techniques cause different levels of spatial awareness [20]. In a study in [21], the researchers examined human eye–head coordination in VR versus physical reality. The results showed that users move their heads more often in VR than in physical reality.

VR is designed to fit in a VE to control the locomotion experiences in a realistic and functional manner. The direction of locomotion is decided by the head-mounted display to point backward, forward or sideway movements [22]. The technology must be customised to master the body movements of a human being and understand the meaning of every command. For instance, the use of finger movements such as pointing, curling, or straightening the fingers helps individuals to carry out real-life experiments using VR. A data glove put on the hands of individual helps to pass commands virtually in a rush against time just like realistic life experiences [22, 23].

To improve the reconstruction of 3D geometry estimation methods based on earlier methods, ML and DL techniques are very tempting and very desirable options [24–26]. To quantify and measure VR sickness during adaptive interactions in the VE, a model based on LSTM was proposed using dynamic information from the normal-state posture signals [27]. Other researchers have discussed the telepresence of the participants from the perspective of behaviour understanding [28]. Abtahi et al. [29] have proposed different methods to enable walking in VR, they found that the experiment will be more immersive when users are at the ground-scale level at the same time increasing the speed of walking to navigate more locations in VR. Physical navigations (including head/body movements) are essential to improve user interaction and engagement in VR applications [30]. Recently, the LSTM provided a good contribution to locomotion prediction in VR [31]. The LSTM has been used to predict the future position after 2.5 s of the current. The research has shown a 65 CM average error for the prediction.

3 Experimental design

To gather the necessary data for our research on mobility in VR, we selected abstract VR painting as the use case for constructing the VE. This choice was made as VR artwork often elicits unpredictable movements from viewers, given each individual's unique perspective and habits when it comes to exploring art. The experimental VE consists of a 3D exhibition room with a large-scale abstract VR painting that is made of tens of thousands of brushstrokes. The VR artwork is placed on one-half of the room, while it is opposite of participants' starting position. Participants can freely move their locations to observe different parts of the artwork from different viewing angles. Participants can also walk into the painting to explore the extensive content inside the painting behind the brushstrokes on the outside.

3.1 Virtual environment

For this research, we designed indoor conditions to construct the VR environment. There are three main elements in the scene: abstract painting, virtual space, and lighting. The abstract painting was used as a core aspect of the environment. Goodyear [32] is a professional VR artist who has created the VR painting for this experiment. She has made several VR artwork exhibitions in public galleries. Goodyear uses Google Tilt Brush to create VR artwork. The painting selected for the experiment consists of several brushstrokes types in a range of colours. The brushstrokes are constricted in a virtual space that allows the participants to walk through. They also take different shapes and styles and have other light conditions. The artist aims to investigate how participants split their attention among these brushstrokes. In previous work, we studied user attention modelling and eye gaze-based community generative art [8, 33]. In this paper, we focus on user interactions related to walking and navigation.

The VE was also designed based on the artist's requirements on how the artwork should be perceived and interacted by the audience besides other environmental settings such as lighting and scaling. The VE consists of a 3D room that has black walls, with a paint pallet as a floor (as shown in Fig. 1) where the painting is placed. The room is scaled to suit the artwork as well the painting pallet is adjusted to suit the walking terrain for the environment. The lighting plays a critical role in imitating the appearance of the artwork. It was customised to produce a bright view over different brushes.

Fig. 1 Virtual and physical environment for the abstract VR painting experiment
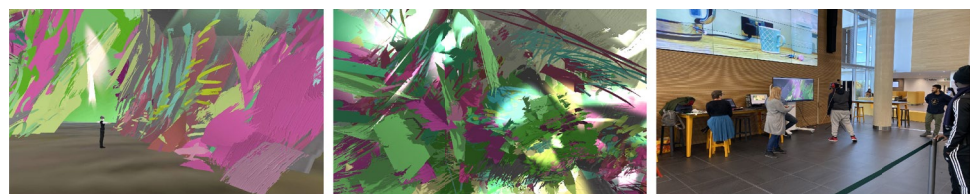
(a) Artwork side view   (b) Artwork side view   (c) Exp. physical space

3.2 Physical space observations

The experiment was arranged in a university public space. Since the experiment aims to study human mobility and behaviour, the authors emphasise having appropriate space and conditions to achieve the research aim. The VIVE Pro Eye comes with two tracking-based stations that could cover a distance of up to 4 ms, which determines the research space as 16 square meters as shown in Fig. 2. We consider the design of physical space to match the virtual space of the artwork. The benefit of this matching is that it elicits a more immersive experience while participants are navigating in the VR.

The experiment space was surrounded by belt barriers that stop participants from travelling beyond the edges of the physical experimental space. Safety measurements were carried out to ensure sufficient precautions for participants as part of the experiment that was carried out during the COVID-19 pandemic.

The design of our experiment aims to incorporate more human factors in order to better understand behaviour. For our study, we chose the HTC VIVE Pro Eye as the primary headset, which offers a high resolution close to 2K and a refresh rate of 90 Hz. The 90 Hz refresh rate is particularly beneficial in reducing simulator sickness, especially since our scene does not contain movable objects [34, 35]. The headset is equipped with an embedded Tobii-based eye tracker for eye-tracking data collection and gazed objects mapping, as well as externally based stations for head orientation and position tracking. Additionally, we incorporated a leap motion device to track hand movements and reactions in the VR headsets. The experimental system also supports the FOVE0 headset, which includes built-in eye-tracking capabilities.

3.3 Participants

Overall, the experiment attracted 35 participants, 20 female and 15 male (Fig. 3). The user information shows that the majority of the participants are aged between 16 and 25 years. More than half of the participants stated that they do not play or rarely play computer games (MD—many times every day, OD—nnce a day, OW—once a week, RL—rarely, NA—not at all). Regarding their experience with VR, 15 had not tried VR before, while 18 had some experience. Only two participants claimed to be very experienced with VR. Similarly, only three participants, who studied fine art, had extensive knowledge of abstract painting, while 18 participants were familiar with this form of artwork (Fig. 3).

During the VR artwork exploration, female participants spent an average of 264 s, which was shorter than the average viewing time of male participants, standing at 276.9 s. Female viewing durations exhibited more variability, as indicated by a standard deviation of 107.7 s, while male viewing time demonstrated less variability with a standard deviation of 53.3 s. Among the participants, the shortest viewing duration recorded was 75.7 s, while the longest duration lasted 548.5 s.
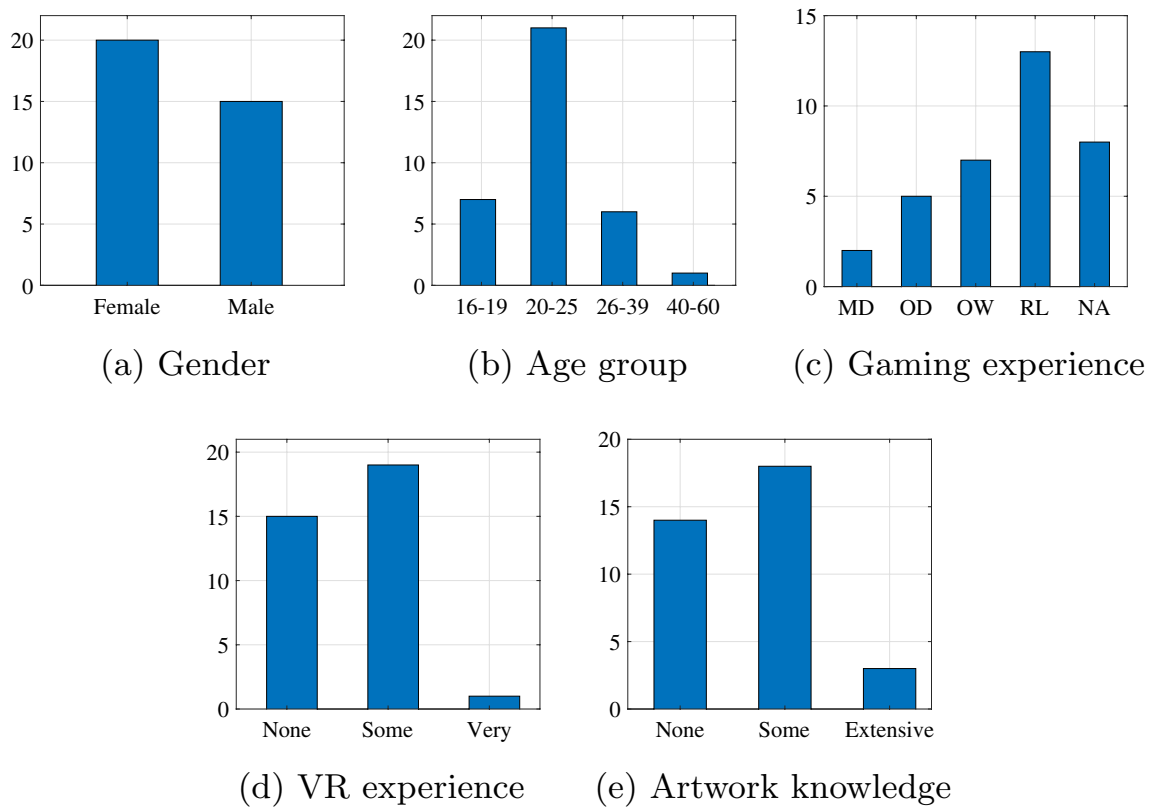
Fig. 2 Floor arrangement for the experiment

(a) Gender (b) Age group (c) Gaming experience

(d) VR experience (e) Artwork knowledge
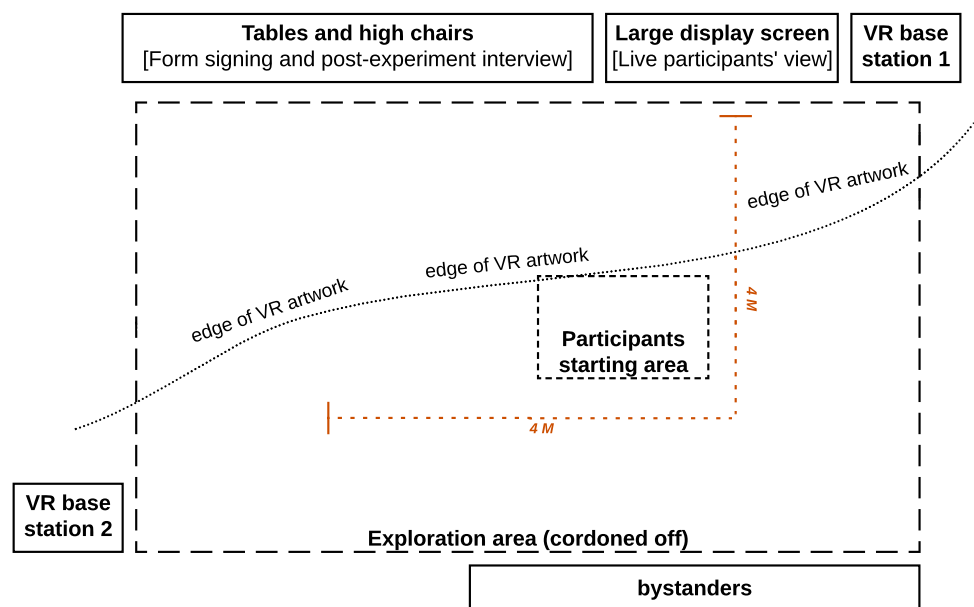
Fig. 3 Participant information

4 Data exploration

This section focuses on mobility data, including head orientation, and position, obtained from the experiment. The data processing approach involved synchronising data and preparing the data for further analysis and modelling.
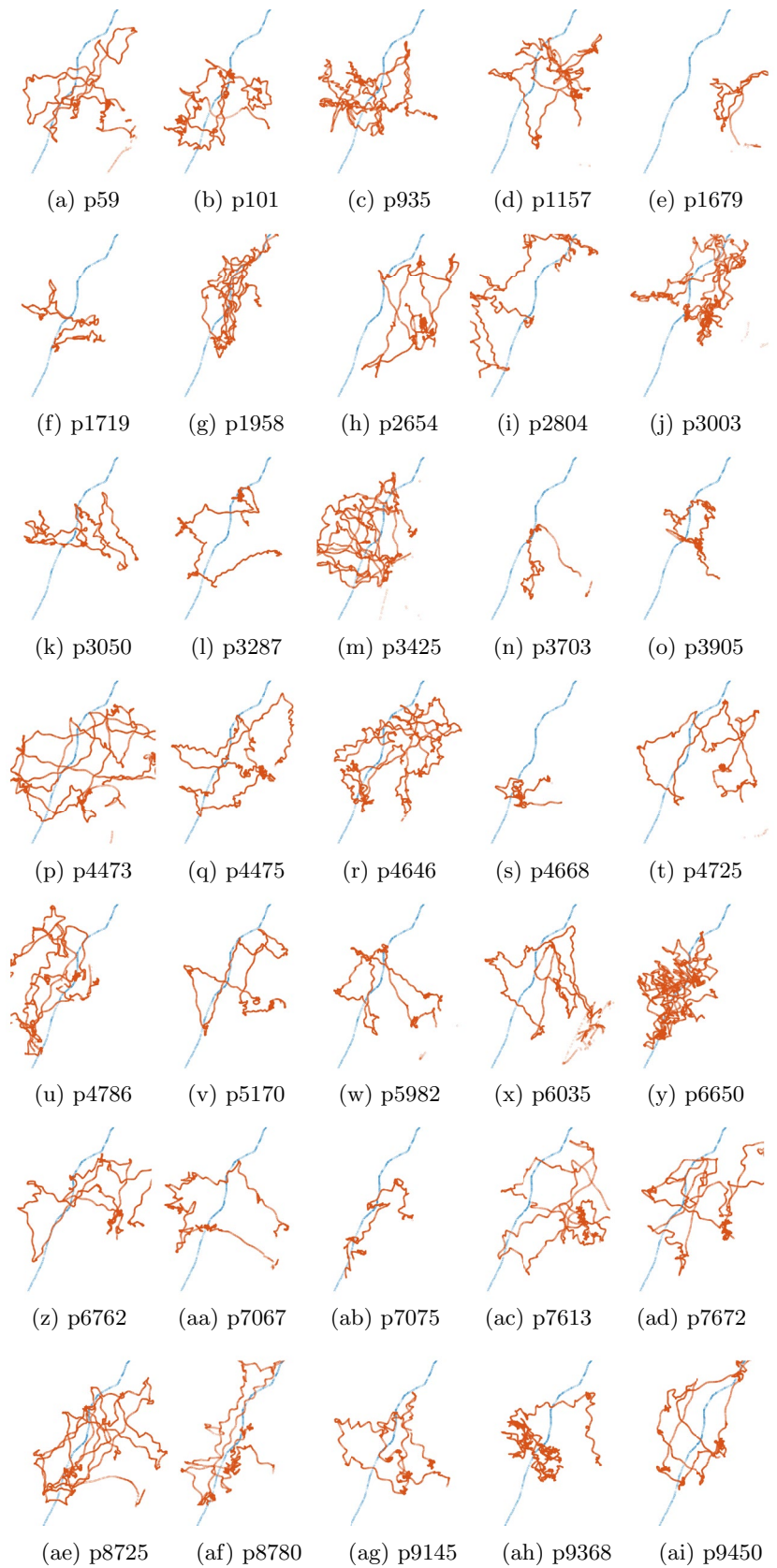
4.1 User movements

The experiment led to a dataset that includes: head orientation, position, eye tracking, and hand tracking [36]. The dataset was gathered using a range of sensors, including headset, position-tracking base, and leap motion. The raw data was collected as follows: head orientation is one of the headset parameters. It represents the head rotation in the VE using four coordinates(x,y,z,w). The player's position is represented in two vectors in the virtual and physical world in (head_x, head_y, head_z) and (player_x, player_y, player_z) consecutively. The Pearson correlation between these vectors is 0.99, which indicates the tracking of the player in the virtual and physical world is highly mapped mobility. Each frame captures these vectors data during the experiment time. This data is also labelled and timestamped to have a recorded journey for the participants.

Figures in 4 show the walk paths from participants. The blue lines mark the edge of the artwork at ground level. The area to the left of a blue line is where the artwork resides, while the area to the right of the blue line is the open space. The orange lines refer to the traces of users' walk inside the VE. All users started the work on the outside of the artwork. The two coordinates used to generate this figure are (head_x, head_z), which represent the head position within the experimental area. It is evident in the figure that participants had different and distinctive walking patterns. Some participants preferred to stay within a small area and mainly viewed the artwork from a distance as how they would behave in a physical art gallery (e.g., p3703 and p1679). Some others enjoyed exploring wider areas by choosing to stay on the outside of the painting and avoiding too much direct virtual contact with any brushstrokes (e.g., p2654, p7613 and p7075). There were also participants who were very adventurous and walked very deep into the artwork (e.g., p3425 and p4786).

The lack of user interactions from some participants reflects a major challenge in designing VR applications in an open VE. It is likely that many participants lacked the necessary knowledge and confidence to navigate the virtual environment effectively, particularly when exploring an unfamiliar setting, such as a new abstract VR painting. We analyzed the raw data from the participants and compared

Fig. 4 Traces of participants' movements (top-down view) (cont.)

(a) p59 (b) p101 (c) p935 (d) p1157 (e) p1679

(f) p1719 (g) p1958 (h) p2654 (i) p2804 (j) p3003

(k) p3050 (l) p3287 (m) p3425 (n) p3703 (o) p3905

(p) p4473 (q) p4475 (r) p4646 (s) p4668 (t) p4725

(u) p4786 (v) p5170 (w) p5982 (x) p6035 (y) p6650

(z) p6762 (aa) p7067 (ab) p7075 (ac) p7613 (ad) p7672

(ae) p8725 (af) p8780 (ag) p9145 (ah) p9368 (ai) p9450

it to the recorded comments and interview questionnaire. Some participants prefer to dive into other objects and attempt to interact by touching brushstrokes. Also, during the walk, participants sometimes lower their body to have a different view of the artwork which is reflected in the changes of their head elevation (head_y). There also appears to be a connection between the change of the head elevation level and the intensity of eye gaze when participants lower or raise their heads to get closer to some objects in the scene. The natural walk generates small elevation waves that can be recognised as a walking pattern. This pattern can help differentiate between walking and head movements while standing.

When participants navigate in the VE and change their location using a free walk, their views of the artwork will also change accordingly. The construction of the environment contains thousands of brushstrokes that fill the virtual space. The brushstrokes are non-collided objects where participants can walk through them. The brushstrokes overfill the VE, where they interfere with each other to create artwork. The interference of the brushstrokes makes the brushes block brushes in behind, so participants are required to walk to view the rest of the artwork. We claim that participants can have multi-views in a busy environment. The reason behind this claim is that a busy environment with many such brushes, colours, and lighting can be rendered or viewed from a single location, So that we assume player mobility can generate different views/sight angles to the environment. Accordingly, this could lead to different levels of attraction for the user. It is possible that examining the sub-scenes within a virtual environment could help explain changes in interactivity or behaviour at different stages of exploration.

4.2 Data processing

Data processing is a crucial step to prepare the gathered dataset for modelling. It involves synchronising data from multiple sensors, mapping actions to interactivity based on game time, and handling missing values, anomalies, bias, and outliers. Human data processing is complex due to the experimental environment. Data cleaning includes removing redundant data, repetitive attributes, and incomplete entries, ensuring the data format is appropriate. Filling techniques are avoided for human activity data.

Removing outliers is challenging due to fluctuating measurements and the absence of a stable reference level for walking data. Sensor-generated outliers can result from factors like bystander interference, participants going out of tracking range, electromagnetic interference, or asynchronous sensor geometry. We have applied statistical moving windows with fixed duration over the data to extract outliers in the data series to improve the quality of the dataset. To

ensure all the data from the sensors have the correct timestamps and are all sequential, we followed mining techniques that verify these data based on a statistical approach. We have considered a threshold to maintain the difference in body velocity. For each position in the data series, we created two windows with the same size of data frames, and considered a static size for the window to ensure the size covered is relevant to the distribution of the data. The edges of the window have no sharp change to the root variance of the window. We calculate the previous and next window velocity in equation 1 and compare the changes of the current participant's position P_c to previous P_{c-1} and next P_{c+1} positions in a range of directions to the w_p, w_N, while we verify the change of velocity based on calculating the root variance for changing the locations over the playtime as in Eq. 2.

$$w_N = \frac{1}{L} \sum_{i=c+1}^{L} \frac{\Delta v_{p_i}}{t}, \qquad w_p = \frac{1}{L} \sum_{i=L}^{c-1} \frac{\Delta v_{p_i}}{t}, \qquad (1)$$

where w_p, w_N are previous and next window consecutively, L refers to window length, c refers to current position index, Δv_{p_i} refers to change of velocity for positions in the window, and t time during Δv_{p_i}.

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (P_i - \mu)^2, \qquad (2)$$

where σ^2 is the square change in mobility of participants, n is the total captured frames per participant. P are positions set where each P has (x, z), μ is the mean of set P.

Following the data pre-processing stage, feature extraction was performed on the raw experimental data. The raw data collected from the experiment are captured per frame (in Fig. 5a). At different time slices, we found different time frames captured from the experiment. This can result in biased results, as the density of data may vary across different experiments. To remove the bias, data samples are aggregated using windows with a duration of 1 s. In each window, we calculate the statistical features for modelling. As a result, we have a dataset with the same data samples across all participants based on a unified timestamp as shown in Fig. 5b.

5 Data modelling

5.1 Clustering walk data

To investigate how participants navigated in VR and how they changed between stand, walking and elevating in different locations, we defined three primary keys behind the mobility of users in VR which are: *(i)* user's personal

Fig. 5 Participants' movements
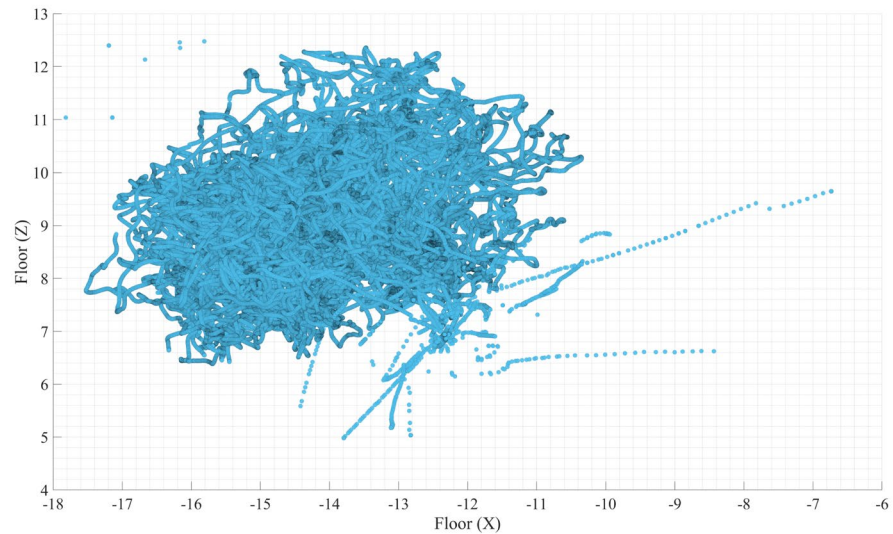
(a) All Participants' movements (x, z) aggregated in the VE
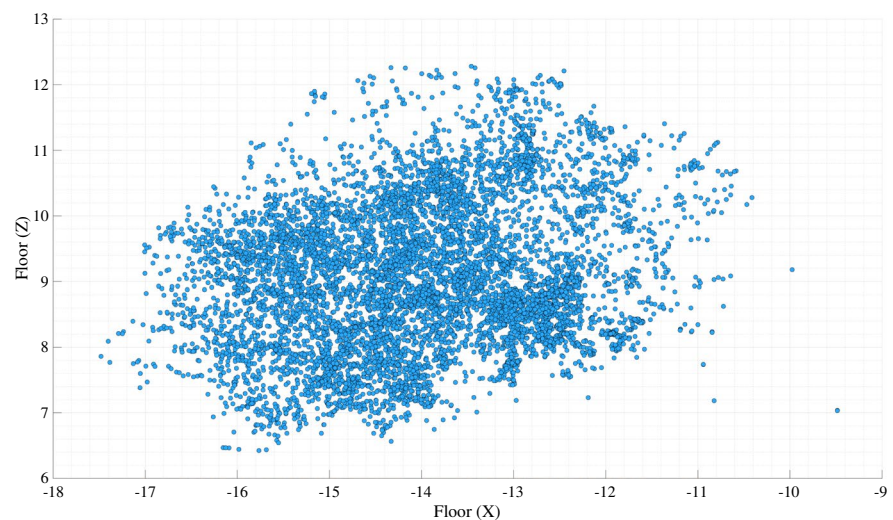
(b) Statistical walk features using one-second fixed window

background such as their previous VR and gaming experience, *(ii)* the VE, and *(iii)* the physical setup. The VE includes the design of virtual elements and environmental characteristics. At the same time, the physical setup includes the headset and the physical space that accommodates the experiment. In this research, we employed these keys in the experiment to develop a dataset that reflects the use of virtual and physical spaces.

Participants moved their locations to explore different parts of the artwork. As a result, any single position in the virtual environment could represent an individual viewpoint with unique brushstrokes and environmental conditions.

Neighbouring positions show similar views as a group (scene), but the views are distinctive between groups. To comprehend the relation between the VE and participants, we mapped the participants' locations in the experiment into clusters using *K*-means clustering to find the most visited scenes in the VE. The decision to use *K*-means was based on the nature of data as we are using Cartesian coordinates (x, z) for grouping.

We experimented with various clustering configurations to investigate the effect of the generated scenes. The process took into account three factors: the distribution of data among these clusters, the distance among the clusters'

centroids, and the diameter of the clusters. The efficiency of the clusters was measured based on the *WCSS* (within-cluster sum of squares) value to test the number of clusters from 1 to 50 clusters, as shown in Fig. 6. A lower *WCSS* value indicates improved clustering results, but normally at the cost of a larger number of clusters. We found that clusters with more than 40 groups suffer from small clustering diameter, resulting in many minimal virtual space and a too fragmented floor space to develop a useful model. For clustering with less than 20 groups, the *WCSS* values are still quite significant and they may not separate distinctive scenes. For clusters where *groups* > 28 *and* < 32, we found a good balance between *WCSS* and the number of clusters as they have an area between $5ft^2$ to $9ft^2$ per cluster.

We have used 30 clusters to maintain the three factors we elected to choose the number of clusters. As shown in Fig. 7, 30 clusters help to reveal more data pattern when building
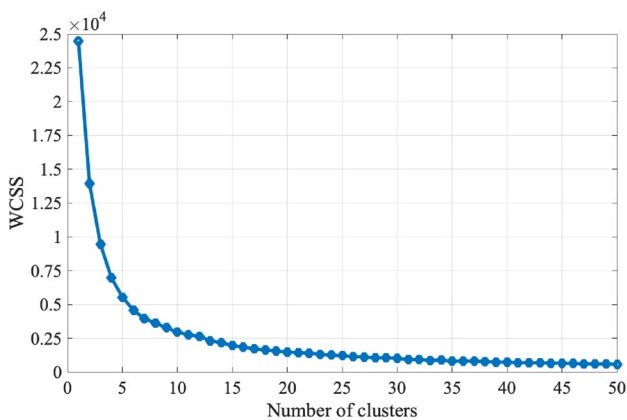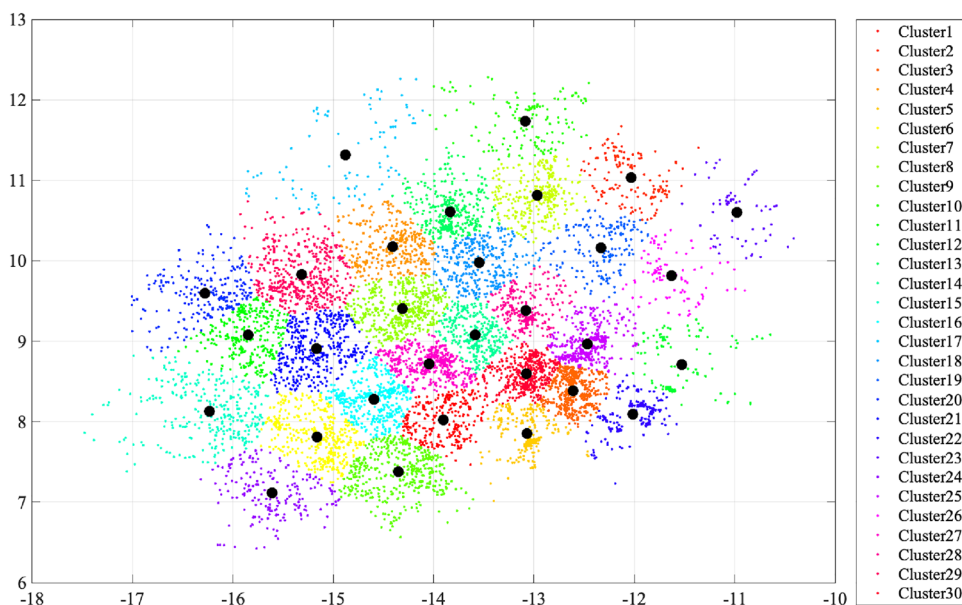
user paths as navigating in a such number of cluster help us to understand user mobility. The format of the current data is the coordinates in the space in addition to cluster number. Each cluster indicates a different view that a participant visited or may visit in the future. We also ensure the clusters cover the entire physical and virtual space. The traces of participants' movements based on clustering-generated views are depicted in Fig. 8.

5.2 Deep learning modelling

The DL modelling aims to discover a common pattern that can capture and simulate user mobility in VR. The advantage of DL is that it allows us to make predictions about complex problems that require discovering hidden patterns and features in the data. We experiment with predicting users' mobility in VR based on users' previous walk steps. Both data from the clustering process and the original data are used to develop different DL-based models to compare their performances. Our data is time series, where each tuple is linked to the next. At the current stage, the data is represented as a flow path drawn from one cluster to another by an arrow. The data has been mined to represent a one- and multi-directions path to easily create subpaths that help models' mobility. After the processing and clustering stage, the dataset has a fixed data density across all time slices. Each data tuple represents a participant's data with a 1-s duration. The data structure consists of spatial coordinates, time, cluster ID, and participant ID.

It is hypothesised that users' previous locations may serve as a determinant in shaping their future movements. Therefore, it is considered a series of historical movements for the prediction of a user's next location. Locations can be defined

Fig. 6 WCSS clusters

Fig. 7 Participants' movements in 30 clusters

(a) p3287

(b) p8780

(c) p7136
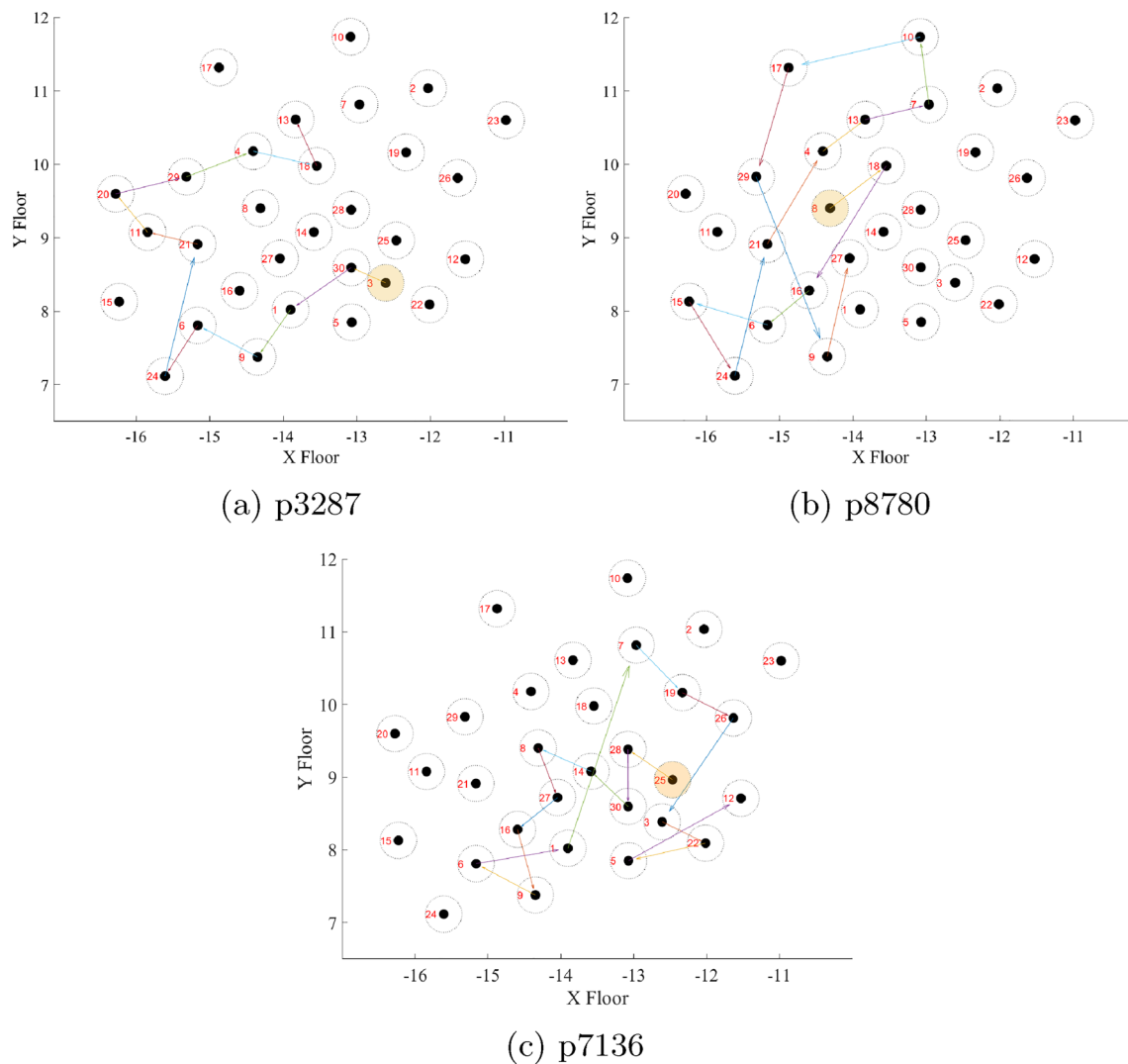
Fig. 8 Traces of participants' movements based on clustering-generated views. Numbers 1–30 represent the ID of the generated cluster/view

as the view or group which has been generated from the *K*-means clustering. The prediction will be based on participants' navigation among the clusters (views) and predict the next potential cluster (view). To achieve the research objective, a dataset of full paths was obtained from 35 participants engaged in VR navigation. Subsequently, a segmentation process was applied to format these paths into smaller sub-paths, facilitating their utilisation as inputs for the DL model. Each sub-path is composed of a sequenced data flow that corresponds to the clusters visited by the respective participant during their VR navigation. Any node on the path is considered as a timestep that a participant has generated from the experiment.

The modelling of subpaths has been tested on various configurations of the numbers of clusters to generate a series of subpaths. Due to the limited paths and views in the data, generating sub-paths with a high number of timesteps

(nodes) leads to a decrease in the total number of generated paths. A lower number of timesteps in generated sub-path can lead to a model with low ML efficiency. Participants' walk paths are constructed using a four timesteps sequence. The participants' mobility was structured into a multiple subpath consisting of four sequential timesteps each, representing the clusters they visited in a specific order. Using the timestamp data, the aim was to predict the next cluster (the fifth cluster) that the participant would likely visit. This predicted cluster served as the expected outcome or future destination.

In the initial phase of the modelling, one-hot encoded data was employed as input to train the DL model (see Fig. 9). The one-hot encoding technique was used to represent the clusters, which served as the input data for the prediction task. Each cluster was represented by a vector with elements equivalent to a cluster ID. Within this vector,

Fig. 9 One-hot encoded progress
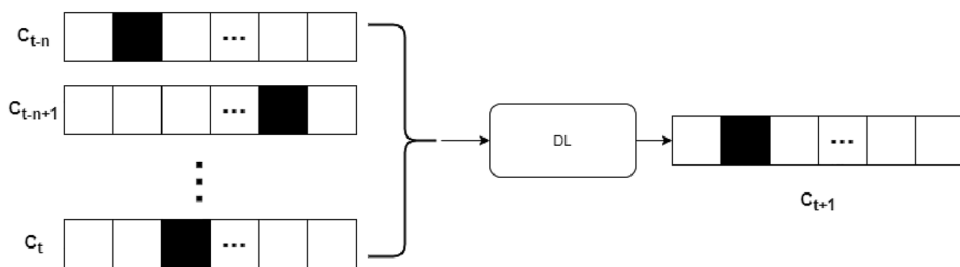
a single element corresponding to the specific cluster was set to one, while all other elements were set to zero. The DL model was trained on this one-hot encoded input to learn patterns and make predictions.

The prediction of patterns was examined using a feed-forward dense network (FDN). An FDN model was constructed to predict patterns within the same dataset. To evaluate the model's performance, various configurations were tested, including different layers and settings for the FDN model. However, despite exploring these different layers and configurations, the results did not show any improvement in the prediction accuracy (Fig. 10). The FDN model was trained using one-hot encoded data as input. Unfortunately, the overall accuracy level remained low, approximately 20%. Various configurations were tested on the model in an attempt to improve results, but no enhancements in prediction were observed.

With the configurations of different numbers of clusters, ranging from 2 to 40, the model's performance did not improve significantly, reaching its peak prediction accuracy at $C_{n=30}$. The utilisation of one-hot encoded inputs had limited influence on modelling mobility, resulting in an average testing accuracy of $\mu = 0.32$. However, the top-K validation accuracy, which considers the top predicted classes, yielded better results with an average accuracy of $\mu = 0.64$

when using $C_{n\pm1}$, $where\ n = 30$. The consideration of top-K accuracy was based on the higher number of classes, where n=30, performing better than the state-of-the-art approach with a prediction probability equivalent to $P(C_{t+1}) = \frac{1}{n}$.

Different techniques are piloted to model the data by using Geo-VR location. Instead of treating clusters as categorical data, their coordinates are used as the input. The Geo-VR location consists of the coordinates *(X, Y)* for the VE floor, which can be mapped to the physical space coordinates. Geo-VR location includes more information of the spatial relationship between clusters. The prediction in this technique is to use these coordinates as an input for the DL and predict the next potential cluster.

Different DL models were developed to assess the learning efficiency and data performance of various prediction techniques. One of these models was based on a recurrent neural network (RNN) called LSTM. The LSTM is a deep learning architecture using Keras sequential API for sequence data (time-series data). It includes multiple LSTM layers with specific parameters, followed by a *"Dense"* output layer with 30 units using the *"softmax"* activation function for classification. The model is compiled with the *"adam"* optimiser and *"categorical_crossentropy"* loss function. Evaluation metrics include *"accuracy"*, *"top_k_ categorical_accuracy"*, and a custom metric *"top3_acc"*

Fig. 10 Accuracy results of the FDN model obtained through tenfold cross-validation

(a) Accuracy using one-hot encoded   (b) Accuracy using Geo-VR
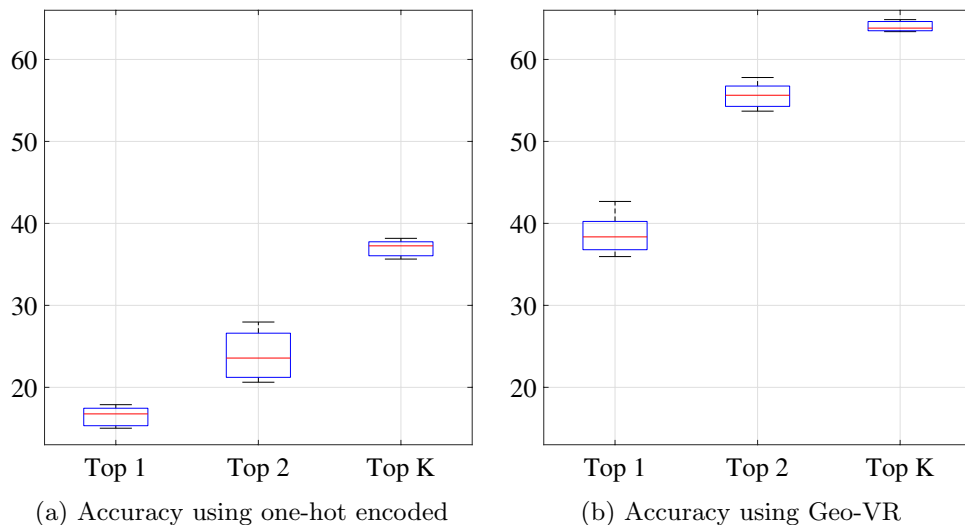
for top-3 accuracy. This architecture is well suited for multi-class classification tasks with 30 output classes and sequence data analysis, specifically designed to handle time-series data (see Fig. 11). Initially, the clusters were treated as categorical data, and an encoding technique known as one-hot encoding was applied. This approach involved representing each cluster as a vector, where each element corresponded to a cluster ID. Specifically, the element corresponding to the respective cluster was set to one, while all other elements were set to zero. The DL model was structured to receive N clusters as input, denoted as $C_{t-n}, C_{t-n+1}, C_{t-n+2}, ..., C_t,$
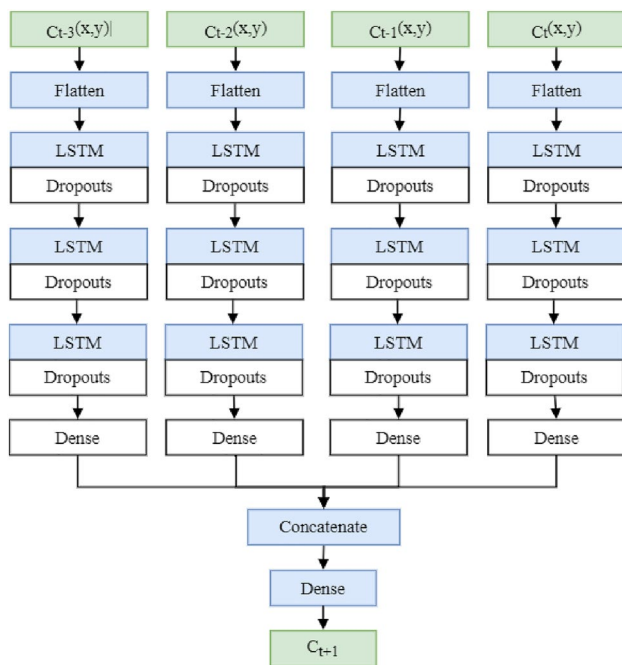

Fig. 11 LSTM model

and predict the next cluster C_{t+1}. The values of 't' and 'n' corresponded to the timestep and the number of sequential clusters or nodes in the participant's path, respectively (Figure 9). The trial of Geo locations has run over the same procedure of generating different numbers of clustering and manipulating model layers. The Geo-Locations have shown a better performance over the one-hot encoded clusters as the testing accuracy average $\mu = 0.66$, with top-K accuracy (0.99), using $C_{n \pm 1}, where\ n = 30$ as shown in Fig. 12.

It is discovered that one of the challenges for accurate prediction is the time step prediction in the correct order. For instance, the DL model may have a very accurate prediction of the next ten movements from a user, but the results can sometimes be out of order when compared with the actual movements done by the user. Two additional approaches have been applied to enhance the time step accuracy: top-K checker and based-nearest destination. The top-K checker is more efficient when calculating the model accuracy as most of the $top(K)$ predictions have a true class in the data label. The top-K checker is more valuable to restrict the options for the potential prediction. It can be used for recommendations on the $top(K)$ predicted classes that can be considered the range of the most likely next views. The based-nearest destination approach is introduced to improve and evaluate the prediction of walk patterns. This technique involves utilising the top neighbour from the prediction vector, where the number of neighbours varies for each cluster, as depicted in Figs. 7 and 8. Implementing the based-nearest destination approach has led to a notable enhancement in the model's performance, with an average accuracy of $\mu = 0.90$.

The impact of the number of K-means clusters on prediction results was investigated. While the clustering analysis resulted in selecting 30 clusters, additional experiments were conducted using a higher number of clusters to predict participants' movements. This led to very complex movement

Fig. 12 Accuracy results of the LSTM model obtained through tenfold cross-validation

(a) Accuracy using One-Hot Encoded

(b) Accuracy using Geo-VR
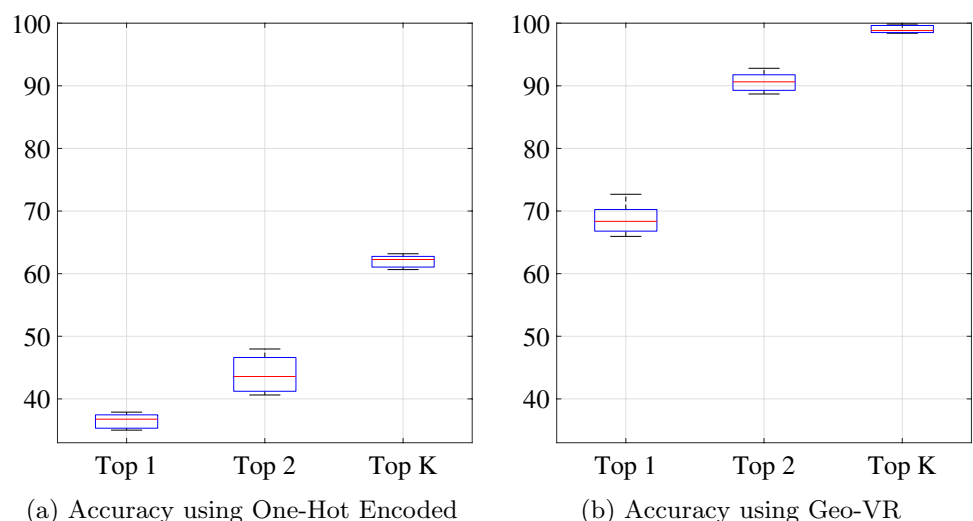
patterns and no improvement to the model performance. Furthermore, gender information was used in combination with the walk data in an attempt to enhance the model based on the hypothesis that there is a correlation between gender and user interaction in VR. However, gender did not show any significant impact on the model performance. This means that male and female participants did not exhibit significantly different mobility patterns while exploring VR paintings.

To assess the accuracy of the prediction model, unseen data was employed for testing purposes. The testing data underwent the same preprocessing methods as the training data, including clustering to ensure their connection with the nearest centroid in the trained data clusters. This ensured that the testing data belonged to one of the clusters previously trained. Subsequently, the testing data was structured into sub-paths of the same length as the trained

data. These sub-paths were then fed into the pre-trained model to obtain the predicted class for each sub-path.

Figure 13 shows the original mobility (left side) of a holdout dataset for a group of participants compared to the predicted mobility (right side) using the DL model. The figure shows clearly that participants consider different navigation patterns while walk in the environment. Some participants chose to take shorter paths when moving within the environment, as shown in Figs. 13b, a, as a result, the area they explored was limited. Some other participants have walked deep into the VR artwork and reached the edges of the tracking area (physical environment) as shown in Fig. 13f, e and h. The model prediction shows exceptional precision results in predicting participants' different travelling patterns in the VE based on the first few steps of their movements.

Fig. 13 Participants' movements (left) vs predicted movements (right)

(a) p1679

(b) p3703

(c) p7067

(d) p3287

(e) p8780

(f) p3003

(g) p4473

(h) p4786
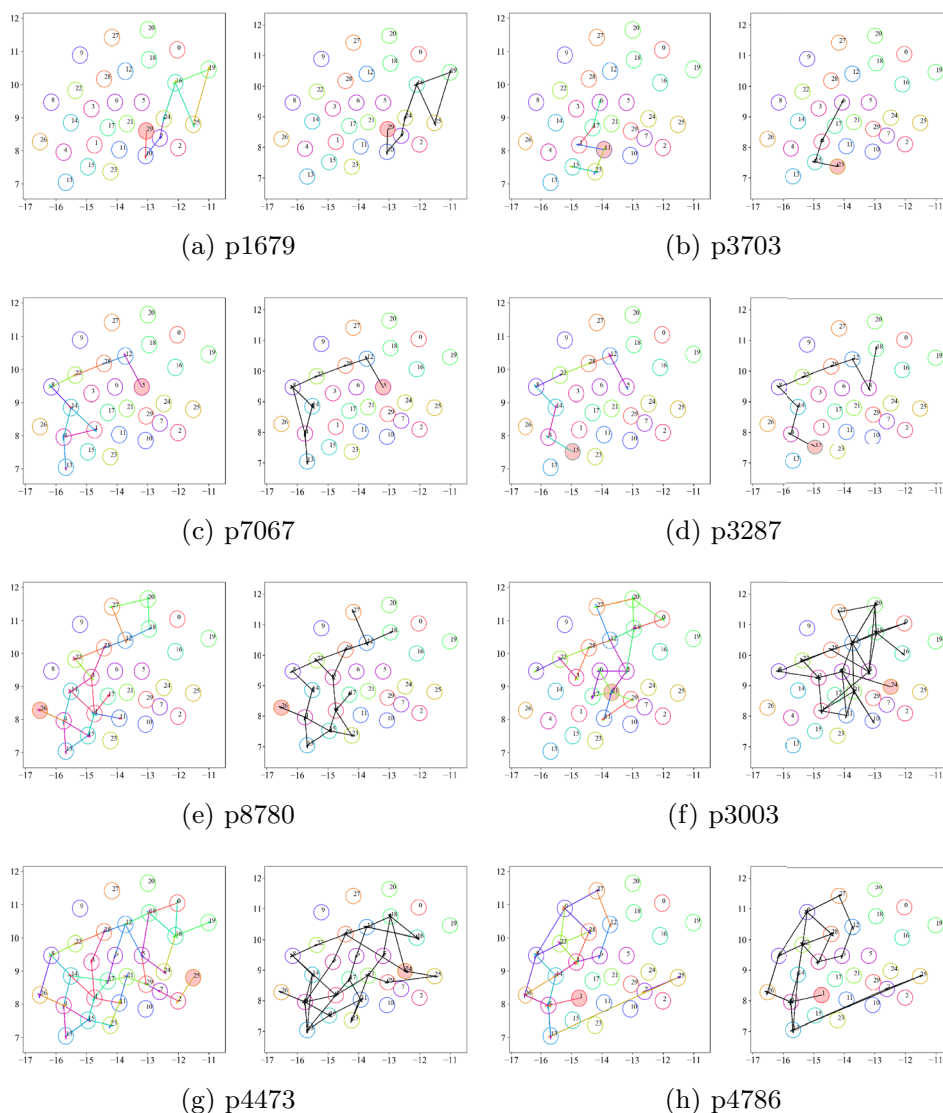
5.3 Discussions

The experimental environment produced a unique dataset that allows us to track user activities and behaviour in a VR environment. The DL models exhibit a high level of accuracy in modelling and predicting user movements in VR environments especially when no clear walking paths and directions were given to the participants. The current modelling consists of different sequential steps that need to be applied in order to obtain the required results. Different models and techniques have been used to improve the prediction. We reached the best performance for the model using clustering, Geo-VR locations and LSTM.

The DL model has its limitations. There is a potential "cold start" issue, similar to that of a recommendation system, that the model cannot draw good inferences when it has not yet received sufficient information. Very short mobility can lead to poor prediction as shown in Fig. 13b, where the participants are not willing to explore more views in the scene then the model may not predict the right next view to be visited. In addition, a participant visited clusters A, B, C, then ending in B will cause a loop in mobility in the input pattern. Looped movements can lead to excessive travels between neighbouring clusters in prediction as shown in 13f.

One of the main use cases of our DL model is a virtual tour guide that recommends areas for exploration to new visitors of a VR exhibition. The recommendation will be based on the first few movements of the visitor and the modelling of user movements from previous visitors or artists. While our results show a high performance in model predictions, the integration of such a model in a VE for recommendation requires additional considerations. Firstly, the model's Top1 suggest for the next location (cluster) may not be immediately next to the user's current location. So the application will need to provide a path for the user to travel or different suggestions can be used from the model's top-K results. Secondly, the model may suggest a path that includes loops (users returning to one of the previously visited locations) based on data from previous users. The loops can be avoided by considering the prediction with the lowest loss to replace the prediction of the model C_{n+1}; Where $C_n \neq C_{n+1}$.

Additionally, it is important to note that the model's generalisability can be assessed due to its consideration of an abstract environment without a given task to model user mobility. The model's ability to use timeseries data to predict user mobility suggests that it could potentially work on other use cases as well. However, for successful application in different environments, certain prerequisites must be met. Specifically, the environmental setup, including the tracking space and real walk locomotion, must be consistent with the original setup used to train the model. Furthermore, the same data processing steps, including the modelling, should be employed to ensure reliable outcomes.

Our work encapsulates a sequence of data processing and modelling steps. Many detailed configurations and ML hyperparameters were tailored for the specific VR environment, physical space, and user interactions. However, the entire process from raw data acquisition to data preprocessing and DL can be automated without human intervention. This is particularly important when the system is deployed to support a public VR exhibition where expert support is limited.

6 Conclusions and future work

In this paper, we conducted a comprehensive study on user free walk mobility in a VR art exhibition, aiming to understand user interactions with the VE and enhance the design of future VR applications. By analyzing complex user movements, we employed a range of ML techniques to define scenes of interest in VR, effectively capturing user mobility patterns. Our LSTM model successfully modelled and predicted participants' movements during VR art encounters, showcasing its strong performance in predicting future navigation movements based on their previous locations. The DL model's capabilities hold significant potential for artists to gain a deeper understanding of audience interactions within the artwork and pave the way for the development of innovative applications, including community-based navigation, virtual art guides, and enriched virtual audience experiences.

Moving forward, we plan to extend our research to explore additional use cases beyond abstract VR painting. Furthermore, we aim to investigate the relationship between users' real-world interactions, eye gaze, and hand gestures, captured during the user experiment. By continuing to study and model user interactions in VR, we can contribute to the advancement of intelligent VR applications that adapt to users' needs and preferences, ultimately enhancing the overall VR user experience.

References

1. Nilsson, N.C., Serafin, S., Steinicke, F., Nordahl, R.: Natural walking in virtual reality: A review. Comput. Entertain. (2018). https://doi.org/10.1145/3180658

2. Langbehn, E., Lubos, P., Steinicke, F.: Evaluation of locomotion techniques for room-scale vr: Joystick, teleportation, and redirected walking. In: Proceedings of the Virtual Reality International Conference - Laval Virtual. VRIC '18. Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3234253.3234291

3. Alghofaili, R., Sawahata, Y., Huang, H., Wang, H.-C., Shiratori, T., Yu, L.-F.: Lost in style: Gaze-driven adaptive aid for vr navigation. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–12 (2019). https://doi.org/10.1145/3290605.3300578

4. Usoh, M., Arthur, K., Whitton, M.C., Bastos, R., Steed, A., Slater, M., Brooks, F.P.: Walking> walking-in-place> flying, in virtual environments. In: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques. SIGGRAPH '99, pp. 359–364. ACM Press/Addison-Wesley Publishing Co., USA (1999). https://doi.org/10.1145/311535.311589

5. Ferracani, A., Pezzatini, D., Bianchini, J., Biscini, G., Del Bimbo, A.: Locomotion by natural gestures for immersive virtual environments. In: Proceedings of the 1st International Workshop on Multimedia Alternate Realities. AltMM '16, pp. 21–24. Association for Computing Machinery, New York, NY, USA (2016). DOIurlhttps://doi.org/10.1145/2983298.2983307

6. Williams, B., Narasimham, G., Rump, B., McNamara, T.P., Carr, T.H., Rieser, J., Bodenheimer, B.: Exploring large virtual environments with an hmd when physical space is limited. In: Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization. APGV '07, pp. 41–48. Association for Computing Machinery, New York, NY, USA (2007). https://doi.org/10.1145/1272582.1272590

7. Steinicke, F., Visell, Y., Campos, J., Lécuyer, A. (eds.): Human Walking in Virtual Environments, 1st edn., p. 402. Springer. https://doi.org/10.1007/978-1-4419-8432-6

8. Mu, M., Dohan, M., Goodyear, A., Hill, G., Johns, C., Mauthe, A.: User attention and behaviour in virtual reality art encounter. Multimed. Tools Appl. (2022). https://doi.org/10.1007/s11042-022-13365-2

9. Marks, B., Thomas, J.: Adoption of virtual reality technology in higher education: An evaluation of five teaching semesters in a purpose-designed laboratory. Educ. Inform. Technol. **27**, 1287–1305 (2022). https://doi.org/10.1007/s10639-021-10653-6

10. Purdy, M.: How the metaverse could change work. Harvard Business Review (2022). https://hbr.org/2022/04/how-the-metaverse-could-change-work

11. Mack, K.: Blortasia: A virtual reality art experience. In: ACM SIGGRAPH 2017 VR Village. SIGGRAPH '17. Association for Computing Machinery, New York, NY, USA (2017). https://doi.org/10.1145/3089269.3089275

12. Hayes, J., Yoo, K.: Virtual reality interactivity in a museum environment. In: Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology. VRST '18. Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3281505.3281620

13. Battisti, F., Di Stefano, C.: Virtual reality meets degas: an immersive framework for art exploration and learning. In: 2018 7th European Workshop on Visual Information Processing (EUVIP), pp. 1–5 (2018). https://doi.org/10.1109/EUVIP.2018.8611753

14. Pfeuffer, K., Geiger, M.J., Prange, S., Mecke, L., Buschek, D., Alt, F.: Behavioural biometrics in vr: Identifying people from body motion and relations in virtual reality. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19. Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3290605.3300340

15. Caserman, P., Garcia-Agundez, A., Konrad, R., Göbel, S., Steinmetz, R.: Real-time body tracking in virtual reality using a vive tracker. Virtual Reality **23**(2), 155–168 (2019). https://doi.org/10.1007/s10055-018-0374-z

16. Lopez, S., Yang, Y., Beltran, K., Kim, S.J., Cruz Hernandez, J., Simran, C., Yang, B., Yuksel, B.F.: Investigating implicit gender bias and embodiment of white males in virtual reality with full body visuomotor synchrony. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–12 (2019). https://doi.org/10.1145/3290605.3300787

17. Abdi Sargezeh, B., Tavakoli, N., Daliri, M.R.: Gender-based eye movement differences in passive indoor picture viewing: an eye-tracking study. Physiol Behav **206**, 43–50 (2019). https://doi.org/10.1016/j.physbeh.2019.03.023

18. Grechkin, T.Y., Plumert, J.M., Kearney, J.K.: Dynamic affordances in embodied interactive systems: the role of display and mode of locomotion. IEEE Trans Vis Comput Graph **20**(4), 596–605 (2014). https://doi.org/10.1109/TVCG.2014.18

19. Ruddle, R.A., Volkova, E., Bülthoff, H.H.: Walking improves your cognitive map in environments that are large-scale and large in extent. ACM Trans. Comput.-Hum. Interact. (2011). https://doi.org/10.1145/1970378.1970384

20. Coomer, N., Bullard, S., Clinton, W., Williams-Sanders, B.: Evaluating the effects of four vr locomotion methods: joystick, arm-cycling, point-tugging, and teleporting. Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3225153.3225175

21. Pfeil, K., Taranta, E.M., Kulshreshth, A., Wisniewski, P., LaViola, J.J.: A comparison of eye-head coordination between virtual and physical realities. In: Proceedings of the 15th ACM Symposium on Applied Perception. SAP '18. Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3225153.3225157

22. Choi, W., Li, L., Satoh, S., Hachimura, K.: Multisensory integration in the virtual hand illusion with active movement. BioMed Res Int (2016). https://doi.org/10.1155/2016/8163098

23. Huang, R., Harris-Adamson, C., Odell, D., Rempel, D.: Design of finger gestures for locomotion in virtual reality. Virtual Reality Intell Hardw **1**(1), 1–9 (2019). https://doi.org/10.3724/SP.J.2096-5796.2018.0007

24. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multiview stereopsis. IEEE Trans Pattern Anal Mach Intell **32**(8), 1362–1376 (2009). https://doi.org/10.1109/TPAMI.2009.161

25. Goesele, M., Snavely, N., Curless, B., Hoppe, H., Seitz, S.M.: Multi-view stereo for community photo collections. In: 2007 IEEE 11th International Conference on Computer Vision, pp. 1–8 (2007). https://doi.org/10.1109/ICCV.2007.4408933. IEEE

26. Ji, M., Gall, J., Zheng, H., Liu, Y., Fang, L.: Surfacenet: an end-to-end 3d neural network for multiview stereopsis. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2307–2315 (2017). https://doi.org/10.1109/ICCV.2017.253

27. Wang, Y., Chardonnet, J.-R., Merienne, F.: Vr sickness prediction for navigation in immersive virtual environments using a deep long short term memory model. In: 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 1874–1881 (2019). https://doi.org/10.1109/VR.2019.8798213. IEEE

28. Rossi, S., Viola, I., Jansen, J., Subramanyam, S., Toni, L., Cesar, P.: Influence of narrative elements on user behaviour in photorealistic social vr. In: Proceedings of the International Workshop on Immersive Mixed and Virtual Environment Systems (MMVE '21). MMVE '21, pp. 1–7. Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3458307.3463371

29. Abtahi, P., Gonzalez-Franco, M., Ofek, E., Steed, A.: I'm a giant: Walking in large virtual environments at high speed gains. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19, pp. 1–13. Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3290605.3300752

30. Ball, R., North, C., Bowman, D.A.: Move to improve: Promoting physical navigation to increase user performance with large displays. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '07, pp. 191–200. Association for Computing Machinery, New York, NY, USA (2007). https://doi.org/10.1145/1240624.1240656

31. Stein, N., Bremer, G., Lappe, M.: Eye tracking-based lstm for locomotion prediction in vr. In: 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 493–503 (2022). https://doi.org/10.1109/VR51125.2022.00069

32. Goodyear, A.: Visual Artist: artist. https://www.alisongoodyear.co.uk/. Accessed 15 Aug 2021

33. Mu, M., Dohan, M.: Community Generated VR Painting Using Eye Gaze, pp. 2765–2767. Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3474085.3478552

34. Wang, J., Shi, R., Xiao, Z., Qin, X., Liang, H.-N.: Effect of render resolution on gameplay experience, performance, and simulator sickness in virtual reality games. Proc. ACM Comput. Graph. Interact. Tech. (2022). https://doi.org/10.1145/3522610

35. Wang, J., Shi, R., Zheng, W., Xie, W., Kao, D., Liang, H.-N.: Effect of frame rate on user experience, performance, and simulator sickness in virtual reality. IEEE Trans Vis Comput Graph **29**(5), 2478–2488 (2023). https://doi.org/10.1109/TVCG.2023.3247057

36. Mu, M.: Explore virtual reality painting with eye gaze. figshare. Dataset (2021). https://doi.org/10.6084/m9.figshare.15090309.v1