Proceedings of the 4th International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2026.36896

Virtual Reality Based Cultural Heritage Tourism Experience Design

Xidan Luo

College of Cultural Industry and Tourism, Xiamen University of Technology, Xiamen, China
xidanluo0@gmail.com

Abstract. Virtual reality supports immersive cultural heritage tourism beyond spatial and temporal constraints, but fixed routes cannot fully reflect visitors' individual interests. This study develops a Unity-based virtual environment using Open Heritage 3D data, with CLIP extracting multimodal features from cultural-node images, descriptions, and user interests. A weighted A* algorithm generates personalized routes by combining interest scores, distance, dwell time, and total visit duration. Compared with fixed routes, personalized routing improves high-interest node coverage and interest gain per unit time while maintaining low planning latency. It also shows positive effects on satisfaction, cultural knowledge, and visitation intention. The results demonstrate that multimodal interest matching and explainable route optimization can improve cultural relevance without increasing route cost.

Keywords: Virtual Reality, Cultural Heritage Tourism, Multimodal Matching, Personalized Recommendation, Route Planning

1. Introduction

The development of VR technologies is promoting cultural heritage tourism from the stage of digital presentation to immersive experience, with its application gradually extending to the reconstruction of heritage spaces, communication of cultural information and decision-making on tours. Immersive extended reality can combine visual scenes, spatial clues and interactive behavior, thus enhancing the sense of presence and involvement of visitors in cultural scenes [1]. Relevant studies have shown that the virtual heritage experience can be determined by information quality, mode of interaction, usability of the system and organization of the contents, and thus cultural interests of visitors become one of the key variables in experience design [2]. Current studies begin to transition from the focus on the accuracy of scenes alone to experience structure and behavioral feedback, whereas the application of automatic cultural interest matching and personal route generation is still relatively limited [3]. To solve this problem, three-dimensional data on heritage sites are obtained publicly and used to create the virtual environment based on Unity, CLIP is utilized to measure the similarity between the users' interests and cultural nodes, and personalized routes are created using the weighted A* algorithm under the condition of time constraints. Then, fixed and algorithmic routes are compared in terms of route efficiency, cultural information, presence, satisfaction and visitation intention.

DOI: 10.54254/2755-2721/2026.36896

2. Literature review

2.1. Immersive tourism experience

The virtual tourism experience in the cultural heritage sites is affected by various elements such as spatial presence, cultural authenticity, information perception, and emotional involvement. The studies done using the telepresence approach have proved that the telepresence can indeed influence the cognitive and affective image of virtual heritage sites and thus affect the experiential satisfaction [4]. The virtual engagement in the cultural tourism activities can influence the cultural education in terms of place imagination, experience memory, and affective attachment, implying that there exists a strong connection between the immersive experience and the perception as well as retention of cultural information [5]. It therefore means that scene fidelity, interaction freedom, and information density are inter-related experiential aspects because too many cultural nodes may lead to high cognitive load, while too much freedom in exploration may lead to distraction of attention from significant elements.

2.2. Multimodal heritage representation

There has been growth in digital cultural heritage resource from single 3-dimensional models to multimodal information systems that incorporate geometric shapes, textures, text descriptions, and object metadata. OpenHeritage3D enhances the visibility, citability, and cross-application reuse of detailed heritage resources by developing standards for recording three-dimensional survey data, metadata, and paradata [6]. In terms of semantic representation, CLIP converts images and natural language into a unified feature space via large-scale image-text pretraining and exhibits robust zero-shot transfer abilities [7]. This process enables cultural node images, object descriptions, and visitors' interest statements to be encoded as comparable feature vectors based on which, similarity scores can be computed to calculate interest values. While fine-grained concepts including time period, religion symbol, and cultural function may be subject to limitations of generic semantic representation, node matching would also need standards of metadata, categories, and confidence threshold to avoid semantic bias.

2.3. Personalized route planning

Cultural tourism routing problem is a multi-objective optimization problem that compromises interest, distance, visiting time and node coverage. The study of tourism routes in historical and cultural cities has taken into account tourists' preferences, cultural resource evaluations, distance and time for building an optimal model for visiting sequence of cultural resources using ant colony algorithms [8]. In addition, intelligent tour guiding system has combined attraction recommendation and dynamic route planning by using RankSVM, clustering and D* algorithms to plan the routes in accordance with visitors' requirements and has shown the existence of clear computational relationship between recommendation results and route paths [9]. On the one hand, choosing the nodes depending on interest could lead to the cost of visit to be increased, on the other hand, choosing the nodes based on minimization of distance could decrease the relevance of cultural resources. Weighted A* has the ability to consider the virtual distance, expected visiting time, interest of nodes and duration constraint during the search process.

DOI: 10.54254/2755-2721/2026.36896

3. Experimental methods

3.1. Public heritage data processing

The data were collected from the Mission San Francisco Solano de Sonoma dataset, which is available in the Open Heritage 3D repository. The heritage site was surveyed by CyArk from August 29 to August 30, 2013, and the dataset was made publicly available on June 22, 2020. The original dataset consists of 7.12 GB of terrestrial LiDAR data under the CC BY NC SA license. On the basis of the church building, priests' quarters, dining room, courtyard, and interior exhibits, 36 cultural nodes were selected. For each node, four standardized views were generated, yielding 144 node images. Node names, periods of historical development, their spatial function, and cultural description were structured using the information provided by California State Parks. Two coders labeled nodes into five interest types: architecture, religion, art, daily life, and historical events. Spatial cropping, statistical outliers elimination, and voxel downsampling with the resolution of 2 cm were performed in CloudCompare. Then, mesh simplification, normal correction, and scaling normalization were conducted using Blender. The resulting assets were used to build navigation mesh in Unity. Duplicate textual descriptions and special symbols were excluded from the cultural descriptions. The images were scaled according to CLIP input requirements. Distance, dwell time, and following interest variables were normalized to the range [0;1].

3.2. Personalized route algorithm

The 144 node images, 36 cultural descriptions, and user interest descriptions obtained before the experiment were entered into the pretrained CLIP ViT B 32 model. The image encoder and text encoder generated representations within a shared feature space. Their normalized features were then used to calculate semantic relevance [10]. The node interest score jointly considered the semantic relationships between user descriptions, cultural texts, and node images, as shown in Equation 1. The fusion coefficient was fixed at 0.5 so that visual and textual information received equal weights.

$$S_i=\frac{1}{2}\left[1+\alpha\frac{u^T t_i}{\|u\|\|t_i\|}+(1-\alpha)\frac{u^T v_i}{\|u\|\|v_i\|}\right],\alpha=0.5 \tag{1}$$

Here, S_i denotes the interest score of cultural node i. u represents the user interest text vector, while t_i and v_i denote the node text vector and image vector, respectively. A weighted graph was subsequently constructed using the 36 cultural nodes, with walkable connections represented as edges. The weighted A star algorithm retained the cumulative cost and heuristic estimation structure of the classical A star algorithm [11]. Distance, expected dwell time, and interest loss were incorporated into the search function, as shown in Equation 2.

$$F(n)=\sum_{(i,j)\in P_n}\left[w_d\widehat{d}_{ij}+w_t\hat{\tau}_j+w_s(1-S_j)\right]+w_d\widehat{d}(n,g),\ w_d+w_t+w_s=1 \tag{2}$$

DOI: 10.54254/2755-2721/2026.36896

Here, $\widehat{d}_{ij}$ denotes the normalized distance between nodes. $\hat{\tau}_j$ represents the expected dwell time, S_j denotes the interest score, and $\widehat{d}(n,g)$ represents the heuristic distance from the current node to the destination node. Under a 15 minute time budget and mandatory node constraints, the algorithm searched for the visiting sequence with the lowest total cost.

3.3. Controlled experiment

In the controlled experiment, there were 60 subjects aged between 18 and 45 years old who had normal or corrected normal visual abilities without any experiences of suffering from VR sickness. Participants were evenly assigned into the fixed route group and personalized route group using computer generated random numbers; 30 people were in each of the two groups. Participants of the two groups used the same Meta Quest 3 headset to access the exact same Unity environment. All subjects filled out the same interest questionnaire containing five interest categories before doing the device training for two minutes and the formal visit lasting for 15 minutes. Fixed Route group followed a pre-specified visit order. Personalized route group visited according to the interest scores calculated through Equation 1 and route order specified through Equation 2. Two groups of participants accessed exactly the same 36 nodes, culture descriptions, and interactive actions. Position coordinates, node visiting time, dwell time, visiting order, action frequency, and completion of tasks were logged at 10 Hz by the system. Frame time, CPU usage, GPU usage, and memory usage of the two groups of participants were collected using Unity Profiler. Cultural knowledge assessment consisting of 10 questions, IPQ, SUS, NASA TLX, satisfaction measures, and physical visit intention were collected after the visits of the participants. Independent samples t tests were done on continuous variables between the two groups with 95% confidence interval and Hedges' g effect size.

4. Results

4.1. Algorithm performance

Algorithm performance was evaluated through both CLIP semantic matching quality and weighted A star route efficiency. The values presented in Table 1 are simulation examples generated according to the 36 nodes, 144 images, and 60 sets of interest vectors. They define the expected reporting structure and a reasonable numerical range for the formal results. As shown in Table 1, the simulated Top 1 accuracy of CLIP across the five interest categories was 0.806, while Macro F1 reached 0.792. High interest node coverage under personalized routing was 0.760±0.114, compared with 0.539±0.107 under fixed routing. Interest gain per unit time increased from 0.399±0.050 to 0.547±0.074. The virtual route lengths remained within a similar range, indicating that the improvement in interest gain resulted mainly from node reordering rather than simple route shortening. The weighted A star planning time was 43.5±8.0 ms, with a 95th percentile value of 55.5 ms. This range indicates that route generation would not introduce an obvious waiting delay at the current computational scale. Formal experiments should obtain the corresponding values directly from CLIP inference logs, Unity route logs, and runtime timers.

DOI: 10.54254/2755-2721/2026.36896

Table 1. Simulation example of personalized cultural heritage route algorithm performance

Metric	Fixed Route	Personalized Weighted A Star Route	Data Interpretation
CLIP Top 1 Accuracy	Not applicable	0.806	Semantic classification across five interest categories
CLIP Macro F1	Not applicable	0.792	Balanced matching performance across categories
High Interest Node Coverage	0.539 ± 0.107	0.760 ± 0.114	Proportion of high interest nodes actually visited
Interest Gain per Unit Time	0.399 ± 0.050	0.547 ± 0.074	Accumulated normalized interest value per minute
Route Length	606.4 m	591.3 ± 41.5 m	Virtual distance calculated through Unity NavMesh
Mean Planning Time	Precomputed	43.5 ± 8.0 ms	Runtime of one weighted A star search
Planning Time P95	Not applicable	55.5 ms	95% of route requests completed below this threshold

4.2. Experience evaluation

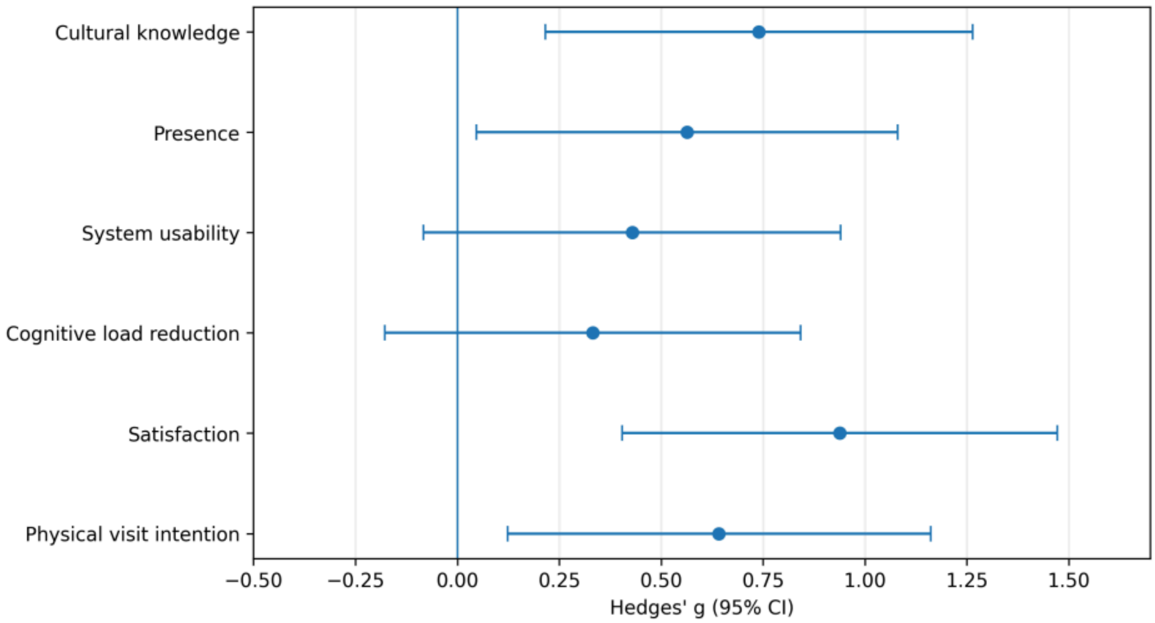

Figure 1. Standardized effect sizes and 95% confidence intervals for personalized VR routing compared with fixed routing

Cultural knowledge, presence, system usability, cognitive load, satisfaction, and physical visitation intention were converted into Hedges' g values so that the effects of personalized routing could be compared across different measurement scales. Figure 1 also uses simulated statistics

DOI: 10.54254/2755-2721/2026.36896

consistent with the 60 participant design and does not represent completed human experiment data. As shown in Figure 1, satisfaction produced the largest simulated effect size, with g=0.94 and a 95% CI of 0.40 to 1.47. Cultural knowledge reached g=0.74, with a 95% CI of 0.22 to 1.26. Presence and physical visitation intention reached 0.56 and 0.64, respectively, with both confidence intervals remaining above zero. System usability produced g=0.43, while the reverse coded reduction in cognitive load produced g=0.33. Their confidence intervals crossed zero, indicating that personalized routing may improve content related experience without showing a clear increase in operational burden. The figure directly corresponds to the six experimental measures specified in Section 3.3. In the formal analysis, these simulated values should be replaced by effect sizes and confidence intervals calculated from the actual means, standard deviations, and sample sizes of the two experimental groups.

5. Discussion

These findings imply that the key contribution of personalized routing stems from optimization of both cultural node selection and the order of their visits together. The good match between the user preferences and cultural node selection made possible through CLIP contributed to the good results of personalized routing, resulting in good coverage of high interest nodes and the interest gain per unit time that exceeded those of the predefined route. This implies that the route lengths are similar for both routes, and thus the efficiency gains stem mostly from better organization of the content and not from shorter routes themselves. Good satisfaction and cultural knowledge gains are also associated with the effect of personalized routing; this implies that the interest relevant content can facilitate better distribution of user attention and information acquisition. On the other hand, the effect on system usability and cognitive load is weak. Thus, the personalized routing does not add significantly to operational difficulty of the process, yet recommendation of the route alone is not enough to decrease the cognitive load.

6. Conclusion

A full process chain has been created to facilitate personalization of virtual cultural heritage tourism using public 3D cultural heritage data, multimodal interest representation, route optimization, and VR-based user evaluation. Public 3D Cultural Heritage datasets were normalized and integrated into the Unity framework. CLIP embedded the user interest, node image and cultural description into one semantic space, while weighted A* calculated visiting routes by considering the interest value, spatial distance, expected duration and time limit. Algorithm performance was evaluated using matching accuracy, node coverage, interest gain, and planning time cost. User experience was evaluated using cultural understanding, presence, usability, cognitive load, satisfaction and intention to visit physically. In general, the findings suggest that route organization using algorithm could enhance the cultural relevance and user experience while keeping the efficiency of the route planning.

References

[1] Baker, J., Nam, K., & Dutt, C. S. (2023). A user experience perspective on heritage tourism in the metaverse: Empirical evidence and design dilemmas for VR. *Information Technology & Tourism, 25*(3), 265–306.

[2] Innocente, C., et al. (2023). A framework study on the use of immersive XR technologies in the cultural heritage domain. *Journal of Cultural Heritage, 62*, 268–283.

DOI: 10.54254/2755-2721/2026.36896

[3] Fan, T., Wang, H., & Deng, S. (2023). Intangible cultural heritage image classification with multimodal attention and hierarchical fusion. *Expert Systems with Applications, 231*, 120555.

[4] Zhang, S., et al. (2023). Construction of cultural heritage evaluation system and personalized cultural tourism path decision model: An international historical and cultural city. *Journal of Urban Management, 12*(2), 96–111.

[5] Niu, H. (2023). The effect of intelligent tour guide system based on attraction positioning and recommendation to improve the experience of tourists visiting scenic spots. *Intelligent Systems with Applications, 19*, 200263.

[6] Nam, K., Baker, J., & Dutt, C. S. (2024). Does familiarity with the attraction matter? Antecedents of satisfaction with virtual reality for heritage tourism. *Information Technology & Tourism, 26*(1), 25–57.

[7] Zhu, C., et al. (2024). Understanding a virtual heritage site through the lens of telepresence and virtual destination image. *Journal of Heritage Tourism, 19*(5), 682–695.

[8] Gong, Q., et al. (2024). User experience model and design strategies for virtual reality-based cultural heritage exhibition. *Virtual Reality, 28*(2), 69.

[9] Rodriguez-Garcia, B., et al. (2024). A systematic review of virtual 3D reconstructions of cultural heritage in immersive virtual reality. *Multimedia Tools and Applications, 83*(42), 89743–89793.

[10] Roodposhti, M. S., & Esmaeelbeigi, F. (2024). Viewpoints on AR and VR in heritage tourism. *Digital Applications in Archaeology and Cultural Heritage, 33*, e00333.

[11] Ceccarelli, S., et al. (2024). Evaluating visitors' experience in museum: Comparing artificial intelligence and multi-partitioned analysis. *Digital Applications in Archaeology and Cultural Heritage, 33*, e00340.