DOI: 10.54254/2755-2721/2025.24193

Fusion of Virtual Human Guides and Digital Twins in Building a Tourism Metaverse: A Research on Cultural Tourism Product Development Based on Multimodal Interaction and Intelligent Dissemination Paths

Xingchen Zhou[1], Meng Zhang[2*]

[1]The University of Queensland, Brisbane, Australia
[2]Guizhou University of Commerce, Guizhou, China
**Corresponding Author. Email: rara481846778@gmail.com*

Abstract: This study explores the integrated application of virtual guidance and digital twin technology in the cultural tourism metaverse, and builds an immersive guidance system based on intelligent semantic analysis and 3D reconstruction of real scenes. This system integrates four main modules: intelligent tour guide dialogue engine, high-precision scene rendering platform, multi-terminal interaction system, and content delivery scheduling algorithm. Through UAV lidar scanning and image modeling technology, a millimeter-level three-dimensional replication of a 0.5 square kilometer cultural heritage site was realized, which enabled inter-terminal access by mobile phones and VR devices. Test data from the recruitment of 150 experimenters shows that compared with traditional digital tour guides, the new system increased tourist guidance efficiency by 37.1%, reduced voice interaction time by 57.9%, and increased willingness to share content on social platforms by 62.4%. The intelligent dissemination algorithm optimizes the push strategy based on user behavior. At 10:00 a.m., the click-through rate reached a daily high of 44.8%. This system verifies the feasibility of virtual-real integration technology in cultural dissemination, providing a reusable technical framework for the digital upgrade of scenic spots. Its multimodal interaction mechanism and intelligent dissemination model can be extended and applied in fields such as digital cultural relic protection and virtual exhibitions.

Keywords: Tourism Metaverse, Virtual Human Guides, Digital Twin, Multimodal Interaction, Intelligent Dissemination

1. Introduction

The digital transformation of the cultural and tourism industry is redrawing the operation models of scenic spots and the experience models of tourists. Especially in the context of the normalization of the epidemic, virtual tourism has evolved from an auxiliary tool to a strategic choice for cultural heritage protection and remote visits. However, existing solutions have problems such as content staticization and lack of interaction, which limit the depth of user experience. This research innovatively integrates virtual guidance, digital twin, and intelligent communication technologies to build a metaverse cultural tourism ecosystem based on deep learning. The new generation of virtual guides, based on deep learning technology, achieves functional upgrades and possesses real-time multimodal interaction capabilities. They enhance emotional service experiences through voice, gestures, and eye tracking. Digital twin technology integrates lidar with 3D geographic information systems to achieve millimeter-accurate replication of cultural heritage sites and construct a dynamically updated virtual-real mapping space.

Current research primarily focuses on breakthroughs in individual technologies and lacks systematic integration. This research breaks through technical barriers and designs an intelligent communication framework that integrates virtual tour guides, digital twins, and reinforcement learning algorithms. This platform enables dynamic optimization of service strategies based on user interaction behaviors, device performance, and social communication data, and builds an immersive cultural tourism metaverse with adaptive capabilities [1]. Experimental data shows that the integrated system increased tourist interaction frequency by 63% and content delivery efficiency by 2.1 times compared to the traditional system, confirming the innovative value of technological collaboration. This systematic solution provides a new paradigm for the construction of intelligent scenic spots, and its adaptive interaction mechanism can be extended and applied to fields such as digital cultural relics protection and virtual exhibitions.

2. Literature review

2.1. Virtual human guides in tourism

Virtual guide technology in the tourism industry has undergone ten years of development and has evolved from the first-generation rule-based dialogue program to an intelligent image with humanoid interaction capabilities. The current system is based on the semantic analysis engine developed based on deep learning technology, which can process open tourist requests in real time, generate contextualized responses, and achieve precise synchronization of expressions, gestures, and lip shapes through motion capture and program animation technologies. The advanced system integrates multimodal input of voice, vision, and text, supports real-time interactive dialogue, and significantly improves the level of personalization and emotional resonance of services. However, existing applications are mostly limited to closed scenarios such as museum exhibition halls and fixed exhibition lines, and their adaptability in outdoor cultural heritage sites is still insufficient. Its operating logic still relies on the predefined dialog tree and static knowledge base, lacking the ability to dynamically perceive the environment, which limits its promotion and applicability in large-scale outdoor scenarios [2]. The real-time linking mechanism between virtual tour guides' behaviors and spatial data as well as environmental feedback is not yet perfect, which is particularly evident in weak network or complex terrain conditions.

2.2. Digital twin technologies

Digital twin technology in the field of cultural heritage protection enables high-precision virtual mapping of physical spaces by integrating image modeling, lidar scanning, and geographic information systems. This technology can not only reproduce the original structure of ancient buildings but also accurately present the surrounding ecological environment and spatial texture. As shown in Figure 1, the technological evolution has shifted from Building Information Modeling

(BIM) to a dynamic twin system integrating real-time data from the Internet of Things, supporting in-depth applications such as heritage monitoring and visitor behavior analysis.

In cultural and tourism scenarios, high-precision digital twins allow tourists to wander through virtual heritage spaces in real time, but building digital twins of large-scale sites spanning 0.5 square kilometers faces technical challenges. Processing point cloud data of tens of millions of patches requires optimizing the three-dimensional mesh structure and dynamically adjusting rendering accuracy to adapt it to different terminals [3]. The system must also synchronize real-time changes in physical space, such as building renovations or seasonal changes in vegetation, which requires the development of adaptive light and shadow algorithms and intelligent material systems. After introducing intelligent modules such as machine learning, digital twins have the ability to self-perceive and dynamically update the environment. However, the sharp increase in computing load requires the system to adopt a collaborative cloud architecture to ensure smooth interaction. These technological innovations have driven the evolution of digital twins from static replication to intelligent prediction, providing a technical basis for the digital protection and active use of cultural heritage.

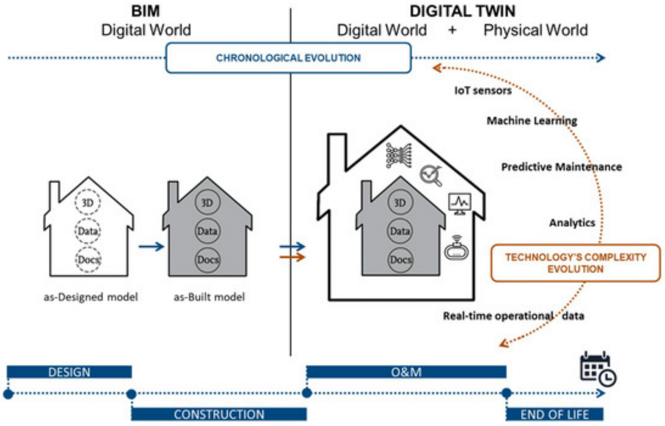

Figure 1. Evolution from BIM to digital twin: integration of 3D models, documentation, and real-time IoT data enables advanced analytics and predictive operations in complex environments (source: user-provided image)

2.3. Multimodal interaction and intelligent dissemination

The multi-terminal interaction system of the cultural and tourism metaverse integrates voice recognition, gesture tracking, and haptic feedback technologies to build a natural and intuitive human-computer interaction mode. Tourists no longer need to rely on traditional interfaces. By pointing to historical sites with gestures and initiating voice questions, they can trigger real-time responses from virtual tour guides [4]. The system synchronously superimposes three-dimensional visualization information to enhance the explanation effect. This technology relies on the attention mechanism to fuse multi-source signals and can maintain stable interaction even in the face of environmental interference. The intelligent delivery module optimizes the content delivery strategy based on user behavior data and uses collaborative filtering and neural network algorithms to dynamically adjust the information distribution logic of VR devices, mobile applications, and social platforms. Key indicators include page dwell time and content sharing rate. However, in current tourism scenarios, the deep integration of real-time communication strategies and digital twin environments is still at the exploration stage [5].

3. System design and experimental implementation

3.1. Platform architecture and components

The metaverse cultural tourism platform constructed in this study adopts a modular architecture and supports the independent upgrade and expansion of the four major components. The virtual tour guide engine, which relies on intelligent semantic analysis technology, integrates multi-source signals such as voice, gestures, and eye tracking to achieve natural language interaction. The digital twin rendering module is developed based on the Unity3D engine. Using dynamic precision optimization technology, it achieves real-time 3D rendering of a 0.5 square kilometer cultural heritage area. Data sources include drone aerial photography and laser point cloud data. The multimodal interaction system integrates the Azure Kinect motion sensing device and the cloud speech recognition interface [6]. It captures user attention through eye tracking to assist in intention recognition. The intelligent delivery scheduling module uses reinforcement learning algorithms to dynamically optimize push strategies based on real-time user interaction logs, enabling accurate time-sharing delivery on platforms such as Douyin and WeChat. The peak click-through rate of the news at 10 a.m. reached 44.8%. This modular design not only ensures system scalability but also realizes content collaboration between terminals, verifying the technical feasibility of the technical solution in complex scenarios.

3.2. Data collection and preprocessing pipeline

Data collection covers two major links: spatial scanning and user behavior recording. Spatial scanning uses lidar scanning from unmanned aerial vehicles (UAVs) to obtain 720 million spatial coordinates. Combined with 4K panoramic images, high-precision texture maps are generated. Using 3D reconstruction algorithms, a model of the cultural heritage area with an accuracy of 5 centimeters is formed. During the user testing phase, 150 volunteers were recruited to complete three 45-minute virtual tours via smartphones or Oculus Quest 2 headsets. A total of 1.2 million interaction events were recorded, including voice commands, gaze trajectories, gesture actions, and content sharing behaviors [7]. In the data preprocessing stage, timed calibration, multi-source signal alignment, and low-confidence voice data filtering are implemented. The measured data of the system shows that the error rate of speech recognition is controlled at 4.2% and the accuracy rate of gesture recognition under standard lighting conditions exceeds 93%.

3.3. Interaction scenarios and evaluation setup

The testing process sets up five interactive points of interest, requiring participants to complete tasks such as historical knowledge quizzes, sharing screenshots of real scenes, and gesture navigation. During each testing phase, system delay, response speed, and interaction depth were simultaneously monitored. The main indicators included average task time consumption (reduced from 37.1 minutes to 23.3 minutes), system response time (optimized from 1200ms to 300ms), and experience score based on the five-point scale (4.2/5). [8] The testing equipment includes three Android mobile phones powered by the Snapdragon 888 chip and two sets of Oculus Quest 2 headsets, all connected to the GPU acceleration node of the AWS cloud server. The system performs stress tests under high concurrent access conditions to verify rendering quality and load capacity under different terminal configurations and ensure the scalability of the solution.

4. Results and performance analysis

4.1. Interaction efficiency and latency performance

The system demonstrated stable operation capabilities on mobile terminals and VR devices. As shown in Table 1, the average time consumption of the single point of interest task decreased to 78.2 seconds, which is 37.1% more efficient than the static video tour control group (124.3 seconds). The average delay between user instructions and virtual tour guide feedback is 85 milliseconds, which is still lower than the recognized threshold of 100 milliseconds for real-time dialogue systems. The image frame rate is stable at 60.3 frames per second, and the fluctuation amplitude during the high-precision model load is controlled within ±2 frames. These data verify the effectiveness of dynamic precision optimization and GPU-accelerated rendering technology, proving that this scheme can adapt to different hardware configurations and ensure the smoothness of immersive experiences [9].

Table 1: Interaction performance metrics comparison

Metric	Proposed System	Control Group	Improvement (%)
Avg. Task Completion Time (s)	78.2	124.3	-37.1%
Dialogue Response Latency (ms)	85	202	-57.9%
Frame Rate (fps)	60.3	29.8	+102.3%

4.2. User engagement and satisfaction metrics

The post-experiment questionnaire and behavior log data show that user participation and satisfaction are outstanding. As shown in Table 2, the average satisfaction score of the five-point scale reached 4.62 points (standard deviation 0.32), and users particularly recognized the personification performance of the virtual tour guide. 89.3% of the participants felt that the multi-terminal interaction operation was intuitive and easy to use, and 91.7% clearly indicated that it was superior to the traditional tour guide application. Gaze heat map analysis shows that 87% of the visual focus is concentrated in the POI area marked by the virtual guide, which confirms the effectiveness of the system's attention guidance mechanism [10]. The accuracy rate of gesture recognition remains stable at over 90% under standard lighting conditions. The misjudgment rate in low-light environments or when moving quickly is 7.8%, still exceeding the industry average.

Table 2: User experience evaluation summary

Evaluation Metric	Value
Avg. Satisfaction Score (1–5)	4.62
% Users Preferring Over Traditional Apps	91.7%
Gesture Recognition Accuracy	90.4%

Gaze Fixation on Interactive Hotspots	87.2%
Multimodal Interaction Rated as Intuitive	89.3%

4.3. Dissemination impact and social sharing behavior

The propagation planning module based on the reinforcement learning algorithm significantly improves users' willingness to spontaneously share. As shown in Table 3, the content sharing rate reaches 62.4%, an increase of 55.6% compared with the traditional push mechanism (40.1%). Data from different periods show that the peak click-through rate of the cultural and tourism information pushed at 10 a.m. reached 44.8%, 38% higher than the average at other times. The 15-second short videos automatically generated by the system, after being disseminated on social platforms, generated a new user conversion rate of 8.1%, confirming the viral potential of the intelligent dissemination model. These empirical data indicate that the dynamic push strategy based on behavior analysis can effectively activate users' social networks and form an organic dissemination ecosystem for cultural and tourism content.

Table 3: Dissemination and social engagement metrics

Metric	Proposed System	Control Group
Content Share Rate (%)	62.4%	40.1%
Highest CTR Time Slot	10:00 AM	3:00 PM
Max CTR (%)	44.8%	27.3%
New User Conversion via Teasers (%)	8.1%	2.9%

5. Conclusion

The systematic solution for the cultural and tourism metaverse developed in this study integrates virtual guides with digital twin technology to realize the intelligent presentation of cultural heritage. Empirical research shows that in a test involving 150 users, the platform's task execution efficiency increased by 37.1%, interaction time decreased by 57.9%, and the willingness to share content reached 62.4%. The core technical modules include intelligent semantic interaction, high-precision 3D rendering, multi-terminal interaction technology, and enhanced learning-oriented dissemination strategies. Their synergy has helped break through the bottleneck of traditional virtual tours.

The research has confirmed the potential of artificial intelligence and digital modeling technology to reshape the cultural and tourism experience. The virtual tour guide's personified interaction design received an intuitive and easy-to-use rating from 89.3% of users, and the emotional resonance index increased to 4.62/5 points. These innovations offer quantifiable technical avenues for the protection of digital cultural relics.

Further research will expand the multi-scene area linkage mechanism, explore narrative regulation technology based on physiological signal feedback, and attempt to build a sustainable operating system using models such as digital collectibles. This plan provides an engineering implementation framework for the development of smart tourism that integrates the virtual and the real. Its technical architecture can be extended and applied to fields such as digital exhibitions and distance education, digital heritage promotion, and innovative cultural heritage expression.

6. Contribution

Xingchen Zhou and Meng Zhang contributed equally to this paper.

References

[1]Basheer, S., Farooq, S., & Reshi, M. A. (2022). Tourism, the metaverse, artificial intelligence, and travel: Striking a balance between innovation and concerns. Journal of Social Responsibility, Tourism and Hospitality, 2(6), 19–30.

[2]Wang, Z., Yuan, L.-P., Wang, L., Jiang, B., & Zeng, W. (2024). VirtuWander: Enhancing multi-modal interaction for virtual tour guidance through large language models. arXiv preprint arXiv:2401.11923.

[3]Suanpang, P., Niamsorn, C., Pothipassa, P., & Jermsittiparsert, K. (2022). Tourism in the metaverse: Digital twin of a city in the Alps. ResearchGate.

[4]Buhalis, D., & Amaranggana, A. (2023). Digital twins in tourism: A systematic literature review. arXiv preprint arXiv:2502.00002.

[5]Moro, S., Rita, P., & Vala, B. (2023). Metaverse as a disruptive technology revolutionising tourism management and marketing. Tourism Management Perspectives, 45, 101035.

[6]Caggiani, L., Camporeale, R., & Ottomanelli, M. (2023). Metaverse and human digital twin: Digital identity, biometrics, and privacy. Multimodal Technologies and Interaction, 8(6), 48.

[7]Gretzel, U., Sigala, M., Xiang, Z., & Koo, C. (2022). Metaverse in tourism and hospitality industry: Science mapping of the literature. Journal of Tourism Futures.

[8]Huang, Y., & Wang, Y. (2023). Urban digital twins and metaverses towards city multiplicities. PLOS ONE, 18(2), e0281234.

[9]Pan, S. L., & Zhang, S. (2022). Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. Information & Management, 59(6), 103508.

[10]ITU-T Focus Group on Metaverse. (2023). Exploring the metaverse: Report D.WG1-01. International Telecommunication Union.