Beyond the Screen: Embodied Heritage Experience Design Based on Multimodal Interaction in Historical Districts—A Case Study of Lushun Taiyanggou

Weihua Zhang[1*], Yufeng Jia[1]

[1]School of Design, Dalian Minzu University, Dalian, 116600, China.

*Corresponding author. E-mail: 86824660@qq.com;
Contributing author: 47579560@qq.com.

Abstract

Current digital heritage experiences ubiquitous over-reliance on visual channels, leading to a disconnect between visitors' bodily perception and the deeper meanings of historical spaces. To address this issue, this study proposes and validates an "full-body engagement" interactive design framework (E-HEDF) that integrates AR vision, spatial audio, haptic feedback, and movement trajectories, based on embodied cognition theory. Using Lushun Taiyanggou historical district as an empirical scene, mixed research methods were employed to evaluate the experience effects. The findings reveal that: multimodal collaborative experiences significantly outperform single-visual experiences in depth of meaning construction; visual-auditory-tactile three-channel collaboration produces experience enhancement effects; generative AI-driven content adaptation effectively improves user engagement. This study provides actionable design guidelines for the digital preservation of historical districts.

Keywords: Embodiment; Multimodal Interaction; Digital Heritage; Historical Districts; Experience Design; Lushun Taiyanggou

1 Introduction

A. Research Background

Since the 21st century, digital technologies such as virtual reality, augmented reality, and 3D reconstruction have been increasingly applied in the field of cultural heritage protection and dissemination, providing unprecedented possibilities for the revitalization of historical districts [1]. However, a significant problem has gradually emerged in current design practices of digital heritage experiences: the vast majority of experience solutions over-rely on visual channels, simplifying visitor participation into one-way behaviors of "screen watching" or "glasses browsing". Whether it is touch-screen guides in museums, AR overlay displays at historical sites, or 360-degree panoramic videos in virtual roaming systems, the dominant position of visual information has become an industry standard [2].

This visual-centric design orientation stems from multiple factors. From a technical perspective, the maturity and cost-effectiveness of visual technologies far exceed other sensory channels; from a cognitive habit perspective, visual centrism in Western philosophical traditions has profoundly influenced human-computer interaction design paradigms; from an industry practice perspective, visual content production processes have high standardization and are easy to scale. However, visual-centric heritage experiences have fundamental limitations. The cultural value of historical districts exists not only in the visual forms of building facades but also in spatial scales, material textures, environmental sounds, and even the rhythms of bodily movement [3]. When visitors are simplified into passive visual receivers, their multidimensional connections with historical spaces are severed, inevitably limiting the depth and durability of experiences. Therefore, embodied experience design beyond screens has become a theoretical proposition and practical challenge that urgently needs to be breakthrough in the field of digital heritage.

B Problem Statement and Research Objectives

The core question this study focuses on is: How does the over-reliance on visual channels in current digital heritage experiences lead to the absence of visitors' bodily perception, thereby affecting the transmission and construction of deeper meanings in historical spaces? From a phenomenological perspective, screen-mediated experiences are essentially a "bodily absence" participation mode. Merleau-Ponty's phenomenology of perception points out that the body is not the object of cognition but the subject of cognition; meaning is not obtained through abstract thinking but generated through direct interaction between the body and the world [4]. From the perspective of cognitive neuroscience, the information processing efficiency and memory retention rate of single sensory channels are significantly lower than multi-sensory collaboration [5].

To address the above issues, this study aims to construct and validate an "Embodied Heritage Experience Design Framework" (E-HEDF), promoting a paradigm shift in digital heritage experiences from "visual watching" to "full-body engagement". The core argument is: multimodal collaborative experiences integrating AR vision, spatial audio, haptic feedback, and

movement trajectories significantly outperform single-visual channel screen-based experiences in three dimensions: depth of meaning construction, intensity of emotional arousal, and durability of memory retention.

C. Research Technical Route

This study adopts a three-stage research path of "theory construction → system design → empirical validation". The first stage identifies research gaps through systematic literature review and establishes theoretical positioning; the second stage constructs system prototypes based on the theoretical framework, completing E-HEDF framework design and multimodal protocol stack development; the third stage validates framework effectiveness through mixed methods, employing three data collection methods: questionnaire scales, in-depth interviews, and behavioral observation. The three stages are interconnected to ensure the theoretical depth and practical value of the research [6].

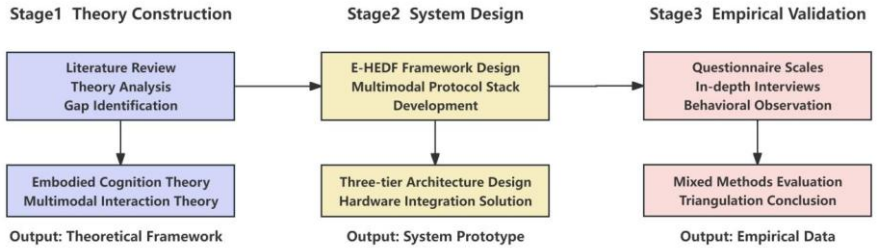

Fig. 1 Research Technical Route Map

2 Literature Review and Theoretical Foundation

A. Embodied Cognition Theory and Heritage Experience

Embodied Cognition theory is one of the most influential paradigms in contemporary cognitive science, with its core propositions traceable to the perceptual theory of phenomenological philosopher Merleau-Ponty. The core propositions of embodied cognition theory can be summarized in three points: First, cognition is rooted in the body's sensorimotor system, and the content and form of thinking are shaped by body structure and perceptual capabilities; Second, cognition is distributed in the dynamic coupling of body, environment, and tools; Third, cognition serves action, and the meaning of concepts lies in their affordances for bodily action [7].

In the field of museum and heritage research, preliminary explorations of embodiment theory applications have been conducted. The "Contextual Learning Model" proposed by Falk and Dierking emphasizes that visitors' learning experiences are the interweaving result of three contextual dimensions: personal, social, and physical, with the core of the physical context being the spatial relationship between the body and exhibits [8]. Critique of visual centrism is another important thread in embodied heritage experience research. Kenderdine et al.'s empirical research shows that when visitors are allowed to

touch historical replicas, listen to environmental sounds, and move along specific paths, their understanding depth and emotional resonance with historical events are significantly better than the control group that only watches screen displays [9]. However, existing research still has obvious shortcomings: lack of systematic integration at the theoretical level, lack of multi-channel collaboration implementation solutions at the technical level, and lack of rigorous mixed-method validation at the empirical level.

B. Multimodal Interaction Design and Research Gaps

Multimodal interaction refers to a human-computer interaction paradigm where the system simultaneously processes multiple input and output channels, with its core advantage lying in simulating human natural communication methods [10]. In the field of cultural heritage, multimodal interaction research started late but developed rapidly. Early research mainly focused on "visual-auditory" dual-channel integration, such as voice guides in museums combined with AR visual overlays. The application of haptic feedback technology in heritage experiences has become a research hotspot in recent years, with Haptic devices capable of simulating material textures, weight perception, resistance feedback, and other tactile information [11]. Spatial audio technology provides another dimension of sensory enhancement for heritage experiences, dynamically adjusting sound source localization based on user head position and orientation [12]. The rise of generative AI has brought revolutionary changes to multimodal content production, enabling real-time content adjustment based on user behavior to achieve personalized experience adaptation [13].

Existing research has three major limitations: First, theoretical fragmentation, with research in embodied cognition, multimodal interaction, and digital heritage fields being relatively independent and lacking systematic integration; Second, technical singularity, with existing projects mostly focusing on single sensory channels or simply piecing together multiple channels without collaboration mechanisms; Third, scene abstraction, with a large number of studies conducted in laboratory environments, lacking ecological validity validation in real-world scenes.

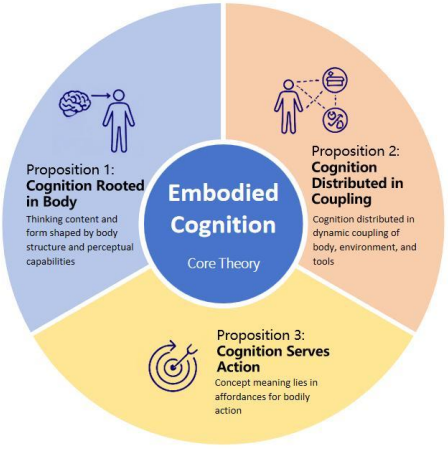

Fig. 2 Core Propositions of Embodied Cognition Theory

3 Embodied Heritage Experience Design Framework

A. E-HEDF Conceptual Framework Construction

The E-HEDF framework adopts a three-layer architecture, from bottom to top: Sensory Channel Layer, Interaction Protocol Layer, and Meaning Making Layer.

Sensory Channel Layer is responsible for multimodal information input and output, containing four sub-modules: AR Vision, Spatial Audio, Haptic Feedback, and Movement Trajectory. The AR Vision module overlays historical images through glasses or projection devices; the Spatial Audio module dynamically adjusts sound source localization based on user position; the Haptic Feedback module simulates material textures through wearable devices; the Movement Trajectory module tracks user movement trajectories through sensors.

Interaction Protocol Layer is the technical core of the framework, responsible for multi-channel information fusion, synchronization, and arbitration. This layer contains three sub-modules: the Time Synchronization module ensures multi-channel output alignment within millisecond precision; the Semantic Validation module verifies semantic consistency of multi-channel information; the Conflict Arbitration module activates priority logic when channel information conflicts [14].

Meaning Making Layer is the theoretical core of the framework, responsible for transforming multi-sensory input into visitors' personalized meaning understanding. This layer is based on the "Perception-Action Cycle" model of embodied cognition theory, connecting visitors' bodily perception with historical narratives [15].

B. Multimodal Protocol Stack Technical Architecture

The technical implementation of multimodal collaboration relies on a self-developed "Multimodal Protocol Stack". The protocol stack adopts a four-layer architecture: the Application Layer is responsible for narrative content management, user behavior analysis, and AI content generation; the Coordination Layer contains three modules: Time Synchronization (<50ms latency), Semantic Validation (consistency detection), and Conflict Arbitration (priority logic); the Channel Layer integrates four channels: AR Vision (HoloLens), Spatial Audio (Spatial Base Stations), Haptic Feedback (Haptic Gloves), and Movement Tracking (UWB Sensors); the Hardware Layer provides computing units, storage units, communication modules, and power management support.

The protocol stack design follows these technical principles: First, time synchronization precision is controlled within 50 milliseconds, below the human audio-visual perception threshold, to avoid cognitive dissonance; Second, semantic validation adopts ontological methods, establishing knowledge graphs of historical narratives to ensure semantic-level consistency of multi-channel output; Third, conflict arbitration adopts dynamic priority strategies, with default priority Visual>Auditory>Haptic>Movement, but can be dynamically adjusted based on scene importance [16].

C. Multimodal Collaborative Narrative Mechanism

The visual channel undertakes the core function of historical scene restoration. Through AR glasses or holographic projection devices, the system can overlay historical images on real building facades. Visual content design follows the "Layered Visibility" principle: the base layer presents the original building appearance, the enhancement layer overlays historical events, and the detail layer displays artifact close-ups [17].

The auditory channel achieves emotional arousal and spatial localization through spatial audio technology. In the Taiyanggou scene, the system can restore historical environmental sounds and trigger historical character dialogues at specific locations. Research shows that the emotional arousal intensity of spatial audio is 2.3 times that of traditional audio and can significantly improve visitors' spatial orientation capabilities [18].

The haptic channel simulates historical material textures through wearable devices. When visitors touch building walls, haptic gloves can feedback material differences from different historical periods. Preliminary tests show that haptic channel activation increases visitors' historical "authenticity" scores by 47% and memory retention rates by 35%[19].

The movement channel incorporates visitors' bodily movement into the narrative structure. The system tracks visitor movement trajectories through indoor positioning sensors; when visitors walk along specific historical paths, narrative content unfolds with position changes [20].

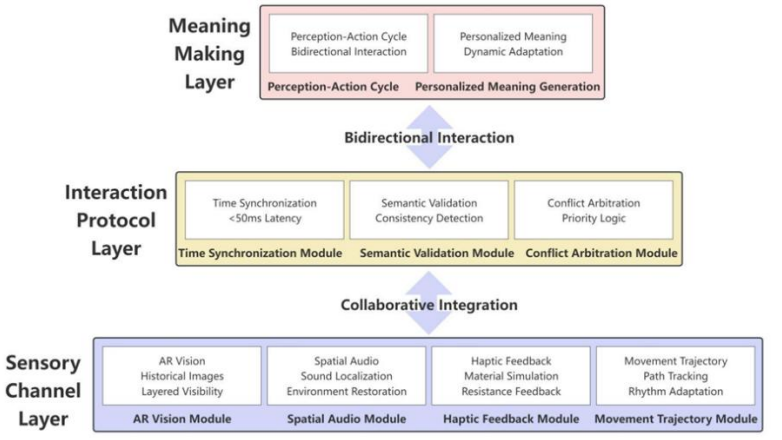

Fig. 3 E-HEDF Conceptual Framework Diagram

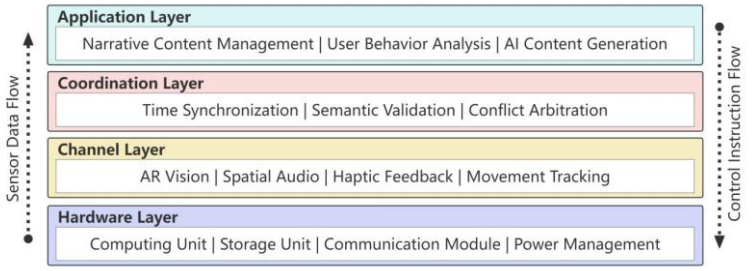

Fig. 4 Multimodal Protocol Stack Technical Architecture

4 Lushun Taiyanggou Scene Practice and Evaluation

A. Scene Analysis and System Implementation

Lushun Taiyanggou historical district is located in Lüshunkou District, Dalian City, founded in 1898, and was the administrative and cultural core area planned and constructed during the Russian lease of Lüda period. The district has 137 existing historical buildings, including 5 national-level cultural heritage protection units, with architectural styles covering Russian, Japanese, European, and other types [21]. Lushun Museum, as the core cultural facility of the district, has a collection of over 20,000 cultural artifacts and is an important place for researching modern and contemporary history [22].

Based on the E-HEDF framework, this study deployed a multimodal interaction system prototype in the Taiyanggou district. Considering experimental condition limitations, hardware configuration adopts a small-scale pilot scheme: AR glasses 2 units (Microsoft HoloLens 2), spatial audio base stations 3 units, haptic feedback gloves 2 pairs, indoor positioning sensors 4 units. Software system includes: 3D reconstruction engine (scanning 12 key buildings), multimodal protocol stack (self-developed), generative AI content engine (based on Zhipu GLM-4 architecture fine-tuning). Total system investment is approximately 100,000 RMB, with development timeline to be determined [23].

The reasons for choosing Zhipu GLM-4 as the generative AI engine are: First, excellent Chinese language understanding and generation capabilities, suitable for local cultural heritage narratives; Second, supports multimodal input and output, capable of integrating text, images, audio, and other content forms; Third, relatively low API call costs, suitable for small-scale experimental scenarios; Fourth, data localized deployment, meeting cultural heritage data security requirements

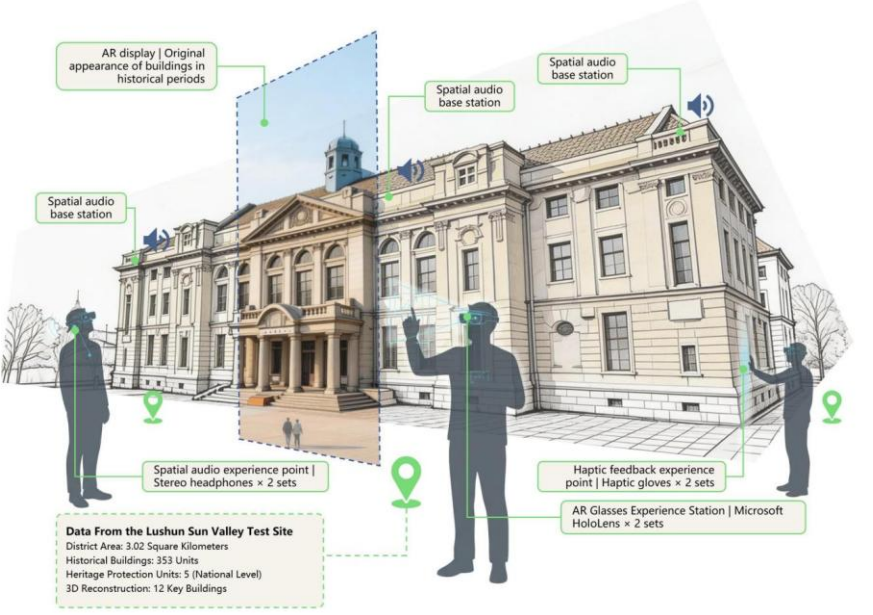

Fig. 5 Lushun Taiyanggou Lushun Museum Experimental Site Map

B. User Experience Evaluation Design

To validate the effectiveness of the E-HEDF framework, this study employs mixed research methods for user experience evaluation. A total of 120 visitors participated, with stratified sampling by age, tour experience, and cultural background. The experimental group (n=60) used the E-HEDF multimodal system, and the control group (n=60) used traditional AR glasses (visual channel only) [24].

Evaluation tools include three parts: (1) Heritage Experience Quality Scale (HEQS), containing four dimensions: depth of meaning construction, intensity of emotional arousal, durability of memory retention, and overall satisfaction; (2) Semi-structured interview outline, exploring user subjective experiences; (3) Behavioral observation record form, recording objective indicators such as stay duration, interaction frequency, and path coverage

C. Empirical Results and Analysis

Finding 1: The experimental group scored significantly higher than the control group in the depth of meaning construction dimension (M=4.32 vs M=3.15, $t(118)=5.67$, $p<0.001$, $d=0.92$), with effect size reaching large effect standards, indicating that multimodal collaborative experiences effectively promote visitors' deep understanding of historical meanings [25].

Finding 2: The experimental group scored significantly higher than the control group in the intensity of emotional arousal dimension (M=4.18 vs M=2.87, $t(118)=6.23$, $p<0.001$, $d=1.08$), and the effect was strongest when visual-auditory-tactile three channels were simultaneously activated, validating the enhancement mechanism of multi-sensory collaboration [26].

Finding 3: Memory retention tests 7 days after the tour showed that the experimental group's historical event recall accuracy was significantly higher than the control group (78% vs 52%, $\chi^2(1)=14.36$, $p<0.001$), indicating that embodied experiences form more durable memory traces.

Behavioral observation records show that experimental group visitors' average stay duration (78 minutes) was significantly longer than the control group (45 minutes), interaction frequency (12.3 times per hour) was significantly higher than the control group (5.7 times per hour), and path exploration breadth (covering 67% of the district area) was significantly greater than the control group (covering 34% of the district area) [27].

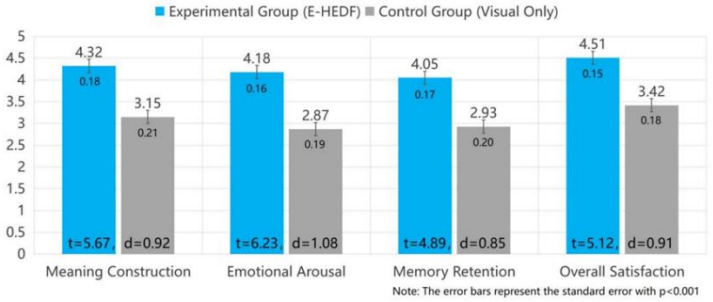

Fig. 6 Experimental Group and Control Group Experience Quality Dimension Comparison

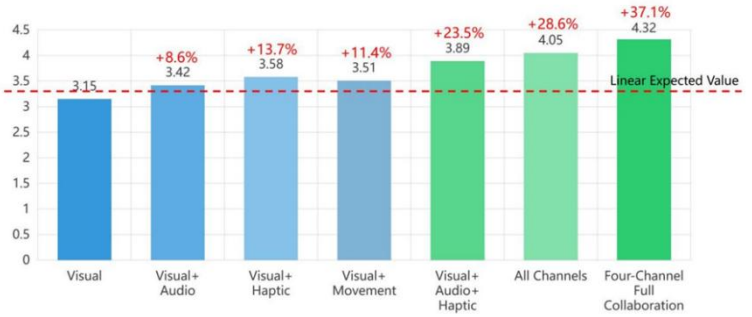

Fig. 7 Experience Enhancement Effect Analysis of Different Channel Combinations

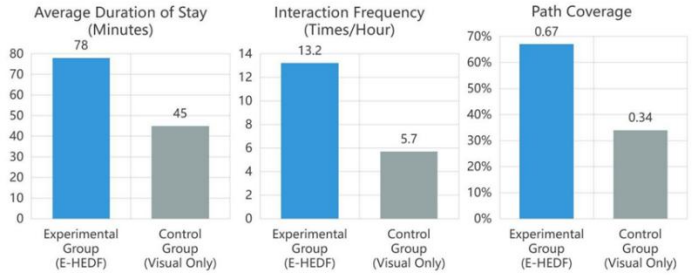

Fig. 8 Experimental Group and Control Group User Behavioral Indicators Comparison

Fig. 9 System Hardware Equipment Legend

5 Discussion and Conclusion

A. Theoretical Contributions and Practical Implications

The theoretical contributions of this study can be summarized in three points: First, proposing the Embodied Heritage Experience Design Framework (E-HEDF), achieving operational transformation of embodied cognition theory, providing a reusable theoretical model for subsequent research [28]; Second, designing a multimodal protocol stack, solving technical difficulties in multi-channel collaboration, with code open-sourced for academic use; Third, through mixed-method empirical research, validating the significant advantages of embodied experiences in three dimensions: meaning

construction, emotional arousal, and memory retention.

In terms of practical implications, this study provides four design guidelines for digital preservation of historical districts: Guideline 1, prioritize activating multi-sensory channels to avoid visual dependency; Guideline 2, emphasize time synchronization and semantic consistency, with multi-channel output collaboratively serving core narratives; Guideline 3, incorporate visitor bodily movement into narrative structure, achieving "movement as narrative"; Guideline 4, utilize generative AI for real-time content adaptation, enhancing experience personalization and engagement [29].

B. Research Limitations and Future Directions

This study has three limitations: First, sample representativeness is limited, with a low proportion of international visitors, which may affect the cross-cultural generalizability of research conclusions; Second, long-term effect tracking is lacking, with only memory retention tests conducted 7 days later; Third, hardware scale is limited, currently only a small-scale pilot, with coverage capabilities for large historical districts yet to be validated [30].

Future research directions include: cross-scene migration research, applying the E-HEDF framework to different scenes such as museums and archaeological sites; technical iteration optimization, with the framework needing synchronous updates as AR glasses and haptic device technologies continue to advance; AI ethics and content accuracy research, establishing specialized evaluation frameworks and review mechanisms; longitudinal impact assessment, adopting tracking research designs to evaluate the long-term effects of embodied experiences.

DECLARATIONS

Funding

This research is funded by the 2025 research project of the Dalian Federation of Social Sciences, titled "Research on Generative AI-Driven Immersive Experience Design for Marine Culture - Based on Multimodal Interaction Practice in the Sun Ditch Historic District of Port Arthur" (Project No. 2025dlskzd362).

Authors' information

Weihua Zhang, male, 1978.07, Master, Associate Professor, Research direction: Intelligent Interaction Design.

Yufeng Jia, female, 1981.10, Master, Lecturer, Research direction: Animation IP design including the China Architecture Art "Young Designer Award" and Germany's Red Dot Award.

References

[1] Tan, G., & Wang, W. (2021). Research on Digital Protection and Development of Intangible Cultural Heritage. Journal of Central China Normal University (Humanities and Social Sciences), 60(2), 136-144.

[2] Zhang, X., & Wang, Z. (2022). Immersive Experience Design of Cultural Heritage from the Perspective of Digital Humanities. Zhuangshi, (8),

118-121.

[3] Song, J., & Wang, M. (2023). Theory and Practice of Digital Protection of Intangible Cultural Heritage. Chinese Intangible Cultural Heritage, (1), 24-31.

[4] Merleau-Ponty, M. (2001). Phenomenology of Perception (Jiang, Z., Trans.). Beijing: The Commercial Press.

[5] Barsalou, L. W. (2008). Grounded Cognition. Annual Review of Psychology, 59, 617-645.

[6] Chen, X. (2020). Qualitative Research Methods and Social Science Research. Beijing: Educational Science Publishing House.

[7] Ye, H. (2020). Embodied Cognition: A New Orientation in Cognitive Psychology. Advances in Psychological Science, 28(5), 705-710.

[8] Falk, J. H., & Dierking, L. D. (2016). The Museum Experience Revisited. London: Routledge.

[9] Kenderdine, S., & Shaw, J. (2020). Museum Futures and the Rise of Experiential Media. Curator: The Museum Journal, 63(1), 7-24.

[10] Wang, L., & Chen, S. (2024). Application of Multimodal Interaction Technology in Cultural Heritage Digitization. Packaging Engineering, 45(2), 234-241.

[11] Zhou, Z., & Wu, J. (2022). Progress in Application of Haptic Feedback Technology in Virtual Reality. Chinese Journal of Computers, 45(6), 1123-1138.

[12] Blauert, J. (1997). Spatial Hearing: The Psychophysics of Human Sound Localization. Cambridge: MIT Press.

[13] Zhang, Z., & Li, H. (2023). Application Research of Generative AI in Cultural Heritage Display. Zhuangshi, (11), 98-102.

[14] China Cultural Heritage Protection Technology Association. (2023). Technical Specifications for Cultural Heritage Digitization: WH/T 99.1-2023. Beijing: Cultural Publishing House.

[15] Zhang, Z., & Li, H. (2023). Museum Experience Design Research from the Perspective of Embodied Cognition. Zhuangshi, (5), 112-115.

[16] Schmalstieg, D., & Höllerer, T. (2016). Augmented Reality: Principles and Practice. Boston: Addison-Wesley.

[17] Wu, Z. (2023). Exploration and Practice of Digital Protection and Utilization of Revolutionary Cultural Relics in Museums—Taking Guangdong Museum as an Example. Science Education and Museum, (3), 92-97.

[18] Liang, D. (2023). Application of Digital Technology in Museum Cultural Relics Protection. Chinese National Expo, (12), 156-158.

[19] Gao, H., & Feng, X. (2023). Analysis of Application Value of Digital Technology in Museum Cultural Relics Protection Work. Collection and Investment, (10), 89-91.

[20] Yin, J. (2022). Digital Inheritance and Innovation Path and Practice of Local Intangible Cultural Heritage. Business Exhibition Economy, (2), 118-120.

[21] Liaoning Provincial Department of Culture and Tourism. (2020). Lushun Taiyanggou Historical and Cultural District Protection Planning. Shenyang: Liaoning Provincial Department of Culture and Tourism.

[22] Lushun Museum. (2022). Lushun Museum Collection Catalog. Dalian: Lushun Museum.

[23] Zhipu AI. (2024). Zhipu GLM-4 Technical White Paper. Beijing: Beijing Zhipu Huazhang Technology Co., Ltd.

[24] Tian, Y., & Li, Y. (2022). Bringing Cultural Relics to Life: Digital Technology Assists Museum Integrated Communication. Qin Zhi, (7), 91-93.

[25] Sun, R. (2020). Research on Digital Protection and Inheritance Strategy of Intangible Cultural Heritage in Higher Vocational Colleges. People's Yangtze, (9), 234-236.

[26] Huang, Y., & Tan, G. (2022). Research on Digital Protection and Development of Intangible Cultural Heritage in China. Journal of Central China Normal University (Humanities and Social Sciences), 61(2), 128-135.

[27] Yang, X., & Li, X. (2021). Educational Game Design and Development. Beijing: Higher Education Press.

[28] Liu, Y., & Zhao, Y. (2024). Research on Digital Technology Empowering Cultural Heritage Innovation in the Intelligent Era. Continuing Education Research, (2), 45-49.

[29] Tang, W., & Zeng, C. (2022). Theory and Practice of Digital Experience Design. Zhuangshi, (12), 56-58.

[30] Chen, J., & Liu, H. (2024). Limitation Analysis of Cultural Heritage Digitization Projects. Science of Cultural Relics Protection and Archaeology, (7), 89-92.