Fluid Cultural Exhibition Grounded in Spatial Production Theory and Generative AI: Interactive Chaozhou Wood-Carving at Bus Stops

Zengqi Ou-yang[1[0009-0007-6548-8528]] XiaoZhang[1[0000-0001-6969-2357]] DanHu[1[0009-0003-3892-8154]] and Li Ou-yang[1[0000-0001-7425-479X]]

[1] The Guangzhou Academy of Fine Arts.Guangdong.510261.CN
oylee@163.com

Corresponding author: Li Ou-yang

Abstract. Background: Museums, as cultural dissemination institutions, have limitations in terms of audience coverage and temporal flexibility in their traditional fixed display mode, making it difficult to meet the fragmented cultural consumption demands of the public. Public waiting spaces such as bus stops, which have the characteristics of prolonged stay time and fragmented attention, provide potential opportunities for informal learning (Informal Learning, IFL). Objective: This study proposes and prototypes a mobile immersive interactive system. This system uses pre-generated generative artificial intelligence (AIGC) content and is intended to be deployed in waiting scenarios. Based on Lefebvre's theory of spatial production, we have constructed a "three-space" interaction framework covering physical, algorithmic, and experiential dimensions. Method/System: Using Chaozhou wood carving patterns as the carrier, the system dynamically presents through the "Tri-state Attention Funnel". Using edge-based visual tracking cameras, the system implements a progressive " Tri-state Attention Funnel " interaction logic: initially, a scattered particle effect is displayed to reduce visual intrusion; when pedestrians are detected, the particles aggregate into the visual images designed by AI to capture attention; if the user stays for more than 10-second threshold, the system will expand detailed textual narratives and element disassembly to support in-depth informal learning. Preliminary Evaluation: We conducted a preliminary field observation in a real bus station scenario. Through the analysis of behavioral coding of one-hour on-site video logs, the conversion efficiency of the attention funnel was evaluated. Key indicators include the conversion rate from "passing by" to "deep participation" at each stage, average stay time, and the embodied behavior performance of users (such as gazing, pointing, etc., natural interaction actions). Ethics and Governance: Prominent notice signs are set up at the scene; following the principle of data minimization, the public video stream is only used for real-time recognition and anonymous behavior recording, and no facial feature data is retained. Contribution: This paper provides a progressive engagement paradigm for mobile exhibitions in public waiting spaces; verifies the engineering feasibility of cultural heritage dissemination based on attention management logic in fragmented scenarios, and proposes a comprehensive solution integrating design strategies and environmental constraints.

Keywords: Media Facades and Media Architectures · Generative AI · Spatial Production · Intangible Cultural Heritage · Interactive Bus Stop · Human–Computer Interaction.

1 Introduce

Museums, as the primary venues for showcasing cultural heritage, have long been constrained by physical walls, facing the 'last mile' dilemma in cultural dissemination. Although the digitalization of museums has been widely discussed, it often merely relocates exhibits from physical displays to online screens, lacking responsiveness to the visitor's immediate physical environment.

To bridge this gap, urban public spaces such as bus stops demonstrate significant potential. These venues serve not merely as transport hubs but as opportunity spaces with inherent potential for informal learning (IFL) [7]. However, in scenarios such as waiting at bus stops, the public's attention tends to be fragmented and passive. Current public facilities predominantly offer functional services, lacking media that stimulate active exploration.

Based on Lefebvre's Spatial Production Theory, space is not a static vessel but a product of social relations. By introducing generative artificial intelligence and human-computer interaction technologies, we can transform static physical waiting spaces into dynamic cultural experiences. This research deconstructs the ornamental characteristics of Chaozhou wood carving patterns—a significant Chinese intangible cultural heritage—to overlay virtual cultural layers in public spaces. Grounded in service design principles, this study proposes a "three-stage attention analysis" mechanism to achieve progressive guidance from environmental cues to deep information transmission, aiming to realize educational value through cultural dissemination in fragmented time.

2 Related Work

2.1 The Potential of Informal Learning in Public Space

In contemporary Western academia, the concept of public space was first introduced by the German political philosopher Hannah Arendt in The Human Condition, wherein she divided the environment of human existence into the public sphere and the private sphere [2]. Arendt emphasized that the public sphere should possess "maximum openness"; this openness provides a field of legitimacy in political terms for informal learning of cultural heritage within the open node of the bus shelter. Public space ought to be a domain where all individuals can be seen and heard. This "visibility" signifies not merely physical openness, but also a shared participation in social life. Traditional urban public space design has long been dominated by functionalism, mechanically partitioning public areas into fixed zones. This fragmented design logic overlooks the multidimensional perceptual needs of people within space [8]. For this study, urban nodes such as bus stops epitomize the public sphere as defined by Arendt—spaces characterized by high mobility, universal accessibility, and diverse social interactions. Museums, long regarded as core public spaces for heritage display and dissemination, often face the "last mile" challenge in cultural transmission due to physical barriers. Despite museums' exceptional expertise, their formal, solemn institutional atmosphere often creates an invisible cultural barrier, making it difficult to reach broad segments of the public who do not actively seek cultural experiences. This "wall effect" has led to cultural heritage becoming somewhat detached from people's daily lives. As relevant research indicates, education should not be confined to formal institutional settings but extended into broader everyday contexts. Hence, shifting focus from everyday contexts to research on informal learning (IFL). Informal learning (IFL) is defined as a spontaneous, unplanned method of acquiring knowledge through daily experiences. According to Hofer (2024), approximately 80% of an individual's learning occurs outside formal institutions such as schools or museums. Informal learning is characterized by its unstructured and unpredictable nature, encompassing experimental, incidental, and random processes [3]. Therefore, the fragmented "passive waiting" time in urban transit scenarios provides ideal "opportunity spaces" to foster unpredictable, experiential, and incidental learning.

2.2 Reconstructing Public Space by Media

Li Minmin (2025) believes that digital design is not only the result of technological innovation, but also an important medium for cultural re-creation and public governance [4]. By reconfiguring the "addition layer" of public spaces through interactive projections, the single functionality of the space can be broken, enhancing the artistry and humanism of urban spaces, thereby achieving service innovation in shared spaces and enabling the new public cultural spaces to maintain continuous innovation in service content, activity forms, and dissemination mechanisms. It emphasizes that cultural services shift from "supply-oriented" to "demand-oriented", leveraging the integration of traditional cultural resources and modern lifestyles, presenting them in a diverse, interactive, and entertaining manner, thereby enabling users to obtain "novelty", "participation", and "cultural satisfaction", and promoting informal learning. [5]

2.3 Generative AI and Cultural Heritage Reconstruction

In recent years, generative artificial intelligence (AIGC) technology has made significant progress, providing a new path for the digital revitalization of cultural heritage from "static symbols" to "flowing narratives". Compared with traditional static exhibitions and conventional digital methods, AIGC technology can leverage algorithms such as style transfer and real-time evolution to efficiently generate dynamic and immersive visual content. This technological paradigm shift not only effectively reduces the production cost of high-quality cultural content, but more importantly, by converting highly complex traditional techniques (such as Chaozhou woodcarving) into intuitive and vivid visual language, it significantly lowers the cognitive threshold for the public to access and understand complex cultural heritage. This enables audiences without professional backgrounds to quickly establish perception and cultural connection with traditional techniques in public spaces.

3 Theoretical Framework

3.1 Lefebvre's theory of spatial production

The spatial production theory of the renowned French sociologist Henri Lefebvre profoundly reveals the dialectical relationship between space and society, and constructs a tripartite dialectical system of "spatial practice - representation of space - representational space". [1] Under the public space context (taking bus stops as an example) that I have constructed, spatial practice (perceived space) points to the daily and empirical material space network. In this article, it is manifested as the physical behaviors of the public during their daily commutes and waiting for trains. Spatial representation (conceived space) represents the conceptual space constructed by designers, planners, engineers, etc. through symbols, logic and order. Representational space (lived space) corresponds to the actual experience of the subject, the space filled with imagination and given new meanings by the users. This is the ultimate goal of space reconstruction, that is, to allow users to break the original functional limitations during the usage process. (see Fig.1)

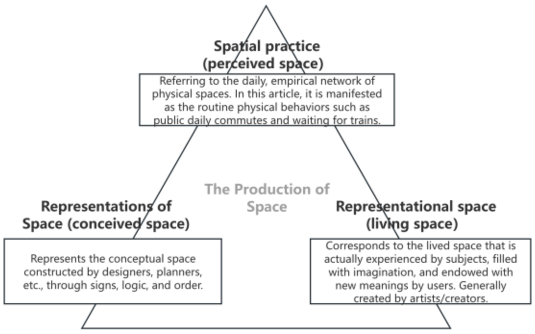

Fig. 1. The tripartite dialectical framework of spatial production theory applied to public waiting spaces.

3.2 Space appropriation strategy

Based on the theory of spatial production, space is not static but a product of social relations. From this perspective, human-computer interaction technology and generative artificial intelligence can be superimposed as an "additional layer" on the existing physical environment, thereby reconfiguring and "appropriating" the architectural environment. "Space appropriation" is not merely a temporary occupation of the physical space, but rather a psychological process of dynamic interaction between humans and the environment. In fragmented and passively waiting public spaces, this study adopts a combination of virtual and real means to superimpose a virtual cultural layer on the physical space. This strategy transforms the originally functionally simple physical waiting space into a dynamic cultural experience and learning domain, enabling informal learning (IFL) to intervene in daily scenarios with embodied and tangible characteristics.

3.3 Mapping between Theory and System Design: Tri-state Attention Funnel

In order to transform the macroscopic spatial theory into a practical interactive engineering, this study conducted a deep structural mapping between Lefebvre's tripartite dialectical system and the core "Tri-state Attention Funnel" of the system: Respecting "Spatial Practice": In the initial state, the system only presents a scattered "particle state", naturally lurking as the ambient background. This low visual intrusion design respects the daily physical space practice of public waiting, avoiding the harsh deprivation of fragmented attention. Intervening "Spatial Representation": When the end-side visual tracking detects pedestrians, the system transforms into "Image State", and the particles aggregate into visual images of Chaozhou wood carvings generated by AIGC. At this time, the algorithm logic actively intervenes, encoding the original physical surface, which was only a traffic node, as a display interface with cultural information carrying capacity, reconstructing the spatial representation system. Creating "Representational Space": When the detection of user staying exceeds the set threshold, the system determines that it has entered the deep participation state, and then unfolds detailed narratives and element disassembly in "Text State". At this time, the original passive waiting of the public is completely reshaped into deep informal learning, and the bus stop platform is "appropriated" as a mobile cultural exhibition hall, achieving a substantive transition from a traffic node to a cultural experience field (i.e., a representational space).

4 System Implementation

4.1 Progressive Interaction

In response to the characteristics of high population mobility and fragmented attention in public transportation station areas, this study has designed and implemented a "three-stage attention funnel" interaction mechanism. This mechanism relies on the computer vision technology of end-side cameras to capture users' spatial behaviors in real time, and precisely divides the interaction process into "particle state", "image state" and "text state".

The displayed content is divided into three stages based on the degree of participation: (see Fig.2)

Particle State.

Trigger Condition. No human skeleton features are detected within the system's field of vision, or the pedestrian merely passes quickly without staying within the effective perception area of the camera.

System Feedback. The screen presents a scattered group of particles moving at a very low rate, simulating the natural Brownian motion of fluids or fine dust. At this point, the system is in a basic standby state, and the visual presentation adopts low contrast and low brightness ambient light effects. This "low visual intrusion" system feedback can make it a digital background (Ambient Background) in the physical space (such as in daily waiting for a vehicle or looking at a mobile phone), providing a gentle environmental hint without disrupting the original spatial practices of the public.

Image State

Trigger Condition. Target enters the interaction distance (1.5-2.5 meters) and the skeleton is recognized stably for over 2 seconds.

System Feedback. The visual system responds immediately, and the scene changes to a particle dispersion state pre-generated by the AIGC tool, aiming to attract users for a deep experience.

Text State. (Deep Informal Learning and Embodied Connection)

Trigger Condition. Continuous dwell time exceeds the 10-second threshold.

System Feedback. The wood carving image evolves from an overall display to detailed local decomposition. The system expands detailed textual narration on the cultural connotations and process dissection. The original passive waiting is reshaped into deep informal learning, successfully creating a representational space.

Fig. 2. Visual representation of the Tri-state Attention Funnel interaction mechanism, showing the transition from Particle State, to Image State, and Text State.

4.2 Technical Architecture: Edge-based Real-time Rendering Pipeline Lefebvre's theory of spatial production

To meet the high-frequency, low-latency demands while strictly adhering to data privacy regulations, this system employs a real-time rendering pipeline built entirely on edge computing (Apple Silicon workstations and TouchDesigner). The input perception layer utilizes edge-side cameras (1280x720) to capture real-time states. The video streams are fed directly into the MediaPipe machine learning framework for real-time local human skeleton landmark detection. Subsequently, through high-level logic evaluation and state-machine selection, the interactive display is rendered via GPU-powered particle systems and transmitted to an on-site projector. Privacy by Design: This architecture adheres to the Data Minimization principle. Skeleton data is extracted frame-by-frame solely in memory; no video streams, facial features, or personal data are stored locally or uploaded to the cloud. End-to-end latency is controlled within 80-120 milliseconds, ensuring a seamless embodied interaction experience. (see Fig.3).

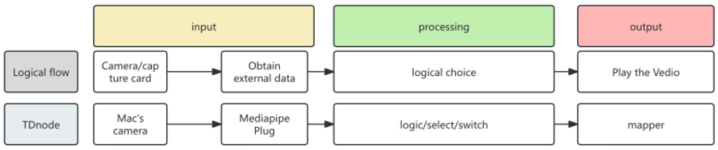

Fig. 3. The edge-based real-time rendering pipeline and technical architecture of the interactive system.

To foster further exploration within the human-computer interaction community regarding informal learning in public spaces, the core interaction logic—the TouchDesigner project files—has been released as open source (GitHub repository: https://github.com/undwait/Chaozhou-Wood-Cave)

4.3 Visual Content Generation and Asset Library Development

Instead of merely digitizing Chaozhou wood carving assets through simple scanning, this study reconstructs its visual symbols using a human-machine collaborative AIGC workflow to meet informal learning needs in public spaces. When designing symbolic patterns, woodcarvers employ various techniques to combine elements, including symbolism, homophonic puns, metonymy, metaphor, and compositional arrangement. [6] Subsequently, this study extracted the decorative motifs of Chaozhou wood carvings by classifying them according to motif type. (see Fig.4)

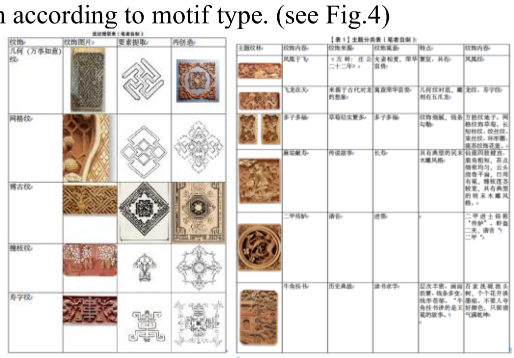

Fig. 4. Characteristic Extraction of Chaozhou Wood Carving Ornaments.

This study developed customized generation strategies tailored to the characteristics of different ornamental types, achieving stable, high-quality outputs:

Geometric and Botanical Motifs. For the regularly arranged ground patterns and borders found in traditional wood carvings, the generation pipeline stably outputs visual images adhering to the principles of "two-way continuity" and "four-way continuity" through symmetry and continuity constraints specified in the prompt. This highly structured pattern not only rigorously adheres to the physical composition principles of authentic wood carvings but also significantly enhances its visual appeal in complex public lighting environments through digitally amplified light and shadow effects.

Architectural and Spatial Elements. Scenes like "pavilions and towers" in Chaozhou wood carvings often feature multi-point perspective and extremely high spatial density, which can reduce visual recognition during digital dimensional reduction. To address this, the research adopted a "cognitive load reduction" strategy when generating such elements. By adjusting prompt weights, the system deliberately downplays overly intricate local physical details, shifting the generation focus toward constructing macro-level visual tension and spatial aesthetics. This ensures these assets rapidly establish visual spectacles for passersby during the "image state (visual capture)" phase of the attention funnel.

Narrative Wood-Carving Panels. These panels carry the highest density of cultural information, centered on the narrative spatial arrangement of characters and scenes. For generating these assets, the system employs a "strong compositional constraint" strategy. The AI strictly preserves classic narrative compositions found in traditional wood carvings (such as zigzag spatial divisions) while digitally optimizing their final presentation medium. This involves using large models to reconstruct surface gold lacquer reflectivity and volumetric lighting effects, endowing traditionally static wood carving panels with immersive qualities suitable for modern projection equipment. (see Fig.5)

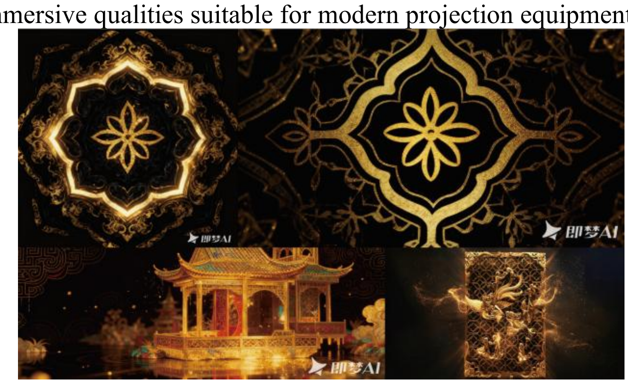

Fig. 5. AIGC-reconstructed narrative wood-carving panels demonstrating enhanced gold lacquer reflectivity and volumetric lighting effects.

Upon entering the TouchDesigner system, Upon entering the TouchDesigner system, the system extracts the RGB channel data from the two-dimensional image and maps it to particle coordinates in Tri-state Attention Funnel.

5 Preliminary Evaluation

5.1 Evaluation Methods: Field Observation and Video Behavioral Coding.

This evaluation employed a non-interventional natural observation approach. Throughout the experiment, hidden cameras continuously recorded the physical space in front of the system to capture the most natural user behaviors. After the experiment concluded, researchers conducted retrospective behavioral coding of the video footage in conjunction with field observation notes. Coding dimensions primarily included: user dwell time within the interaction zone, the highest-triggered system state (particle/image/text), and users' natural physical and social interaction behaviors.

During the effective observation period, the system captured approximately 33 natural foot traffic instances passing through the scene. After excluding one young child sample deemed too young to provide valid cognitive feedback, the effective sample size was N=32. Analysis of these 32 users' behavioral data revealed the following funnel conversion distribution characteristics:

Stage 1 (Environmental Glance). 22 users (approximately 68.75%) had dwell times under 5 seconds. This group primarily exhibited rapid commuting behavior. The system successfully maintained a low-intrusive "particle-state" environmental background, avoiding visual disruption.

Stage 2 (Visual Capture). 3 adult users (approximately 9.38%) were drawn to aggregated woodcarving images and paused, with an average dwell time of about 15 seconds. They triggered the "image-state" but did not reach the dwell threshold for deep learning engagement.

Stage 3 (Deep Informal Learning). Seven users (five adults and two school-aged children, 21.87%) maintained continuous engagement exceeding 30 seconds—far surpassing the system's 10-second trigger threshold. They successfully activated the "text-state" narrative and demonstrated focused reading behavior.

Data indicates that in fragmented public spaces, the system successfully converted over one-fifth (21.87%) of passersby into deeply engaged cultural participants, preliminarily validating the engineering efficacy of this progressive interaction paradigm in capturing and sustaining attention.

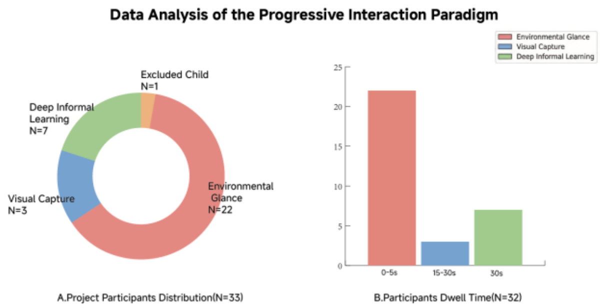

Fig. 6. Data analysis of the progressive interaction paradigm.

5.2 User Behavior Patterns

Beyond quantitative dwell time, video behavior coding revealed several typical human-machine/human-human interaction patterns in public spaces:

Pattern One: Child-driven Intergenerational Engagement. Multiple instances were observed where children actively explored and subsequently drew accompanying adults into shared participation. This confirms the potential of public environments as informal "opportunity spaces" for learning, where visually spectacular interactive installations effectively stimulate children's curiosity, transforming it into family-based collaborative learning.

Pattern two: The Honeypot Effect. Video footage shows that when users engage in "deep learning (text-based)" states, their embodied participation behaviors (e.g., prolonged staring, approaching) serve as strong social signals, drawing bystanders who were about to leave to gather and watch. This "people-attracting-people" clustering phenomenon significantly amplifies the cultural dissemination radius of single-point digital interactive installations in public spaces.

Pattern Three: System-Cognition Mismatch. Behavioral coding revealed a notable "False Positive" case. One user met the dwell time threshold to trigger the text mode within the interaction zone, yet their body language (e.g., looking around, frowning) indicated confusion rather than actual deep reading. This demonstrates that while "dwell time" based on pure visual tracking serves as an effective engineering proxy metric for attention, it cannot fully equate to the user's internal cognitive state. This highlights a potential optimization direction for future system iterations: multimodal intent recognition (e.g., incorporating eye-tracking interventions).

6 Discussion and Limitation

Based on our development and field deployment experience with this system, we propose the following three insights for future HCI design targeting urban public spaces:

Implicit Awakening and Explicit Capture in Dynamic Environments. Digital designs should incorporate "stealth" states to blend into the environment, using algorithms to create visual spectacles at opportune moments to balance aesthetics with information dissemination.

Leveraging the "Honeypot Effect" to Amplify Social Gravity. Public interactions are contagious social performances. Future designs should incorporate this "Performer-Spectator" dynamic to broaden dissemination.

Privacy by Design. Employing pure edge computing with instant data destruction establishes a mandatory compliance model for visual interaction systems in public spaces.

7 Conclusion

Future work will incorporate gaze tracking or multimodal intent recognition to more precisely capture user attention states. Additionally, we plan to introduce subjective experience scales (e.g., perceived value scales) and explore extending AIGC models to enable automated generation and interactive storytelling across broader intangible cultural heritage domains.

Acknowledgments. We gratefully acknowledge the financial support from the following pro-grams: 2024 Guang-dong Provincial Graduate Education Innovation Plan

(2024SFKC_054), 2025 Guangdong Pro-vincial Graduate Demonstration Course Construction Project "Visual Design and Research of Traditional Culture" (2025KCJS_070)

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Lefebvre, H.: The Production of Space. Blackwell, Oxford (1991)
2. Arendt, H.: The Human Condition. Shanghai People's Publishing House, Shanghai (2017)
3. Hofer, L.: Opportunity Spaces for Children's Informal Learning in Public Environments. In: Proceedings of the Eighteenth International Conference on Tangible, Embedded, and Embodied Interaction, pp. 1-5. ACM, Cork, Ireland (2024)
4. Li, M.: How to Intervene? A Study on the Reshaping of Urban Public Spaces Through Digital Design (n.d.)
5. Zhang Haojue, Li Yawen, Li Si. Measuring Users' Perceived Value in New Public Cultural Spaces [J]. Library and Information Service, 2026, 70(4): 71-83.
6. Chen, L.: The Application of Traditional Chinese Patterns in Wood Carving Craft Design. New Horizon (11), 58-60 (2022)
7. Liu, Y., Xu, H., Zhou, Y.: Digital Innovation based on the Perspective of Public Art Drives the Improvement of the Quality of Urban Public Space 30(1), 1-5 (2023)
8. Huang, W., Yang, Q.: The Virtual Construction of Urban Public Spaces by Interactive Art. Contemporary Art (9), 77-79 (2025)